2025 공동 학술대회

Barun ICT Research Conference

Agentic AI for Responsible Innovation and Value Up

한국지식경영학회 추계학술대회

Agentic AI 시대, 안전한 정보와 지식

ISACA Korea Conference

Agentic AI 시대, Digital Trust와 IT 거버넌스

2025. 11. 13.(목) 09:00 ~ 17:30

연세대학교 백양누리 그랜드볼룸 및 컨퍼런스홀

연세대학교 YONSEI UNIVERSITY | 바른 ICT연구소 Barun ICT Research Center | 연세대학교 정보대학원 GRADUATE SCHOOL OF INFORMATION | The Knowledge Management Society of Korea (사)한국지식경영학회 | SMU SINGAPORE MANAGEMENT UNIVERSITY Centre for Digital Law | ISACA Korea Chapter

Acknowledgements

2025 공동 학술대회 조직위원회는
다음의 지원 및 후원자에게 감사의 말씀을 전합니다.

U+
FIS 한국재정정보원
WOORIINIS (주)우리아이아이에스
ITCEN ENTEC
SEEDGEN.
(주)씨에이에스 (C·A·S) Computer Assurance Services
TB (주)투비콤
ZETACUBE™ The Solution leader for ZiB Storage
WIT (주)우리아이티 Leading company in IT convergence
SOMANSA
Selim 세림TSG
NEOBRIX
나날 SMI 나날에스엠아이(주) NaNal System Maintenance & Integration
GIT 굿모닝아이텍(주)
SK telecom
FROM100
BK FOUR 21 연세대학교 정보대학원 AI-빅데이터 기반 초스마트 사회 구현을 선도하는 교육연구단
NRF 한국연구재단 National Research Foundation of Korea
GITSN
First 퍼스트치과
도서출판 청람
초아출판사

한국지식경영학회
2025 추계학술대회 조직위원

학술대회장
김범수(연세대학교 바른ICT연구소)

조직위원장
고준(전남대학교)
이호(한국기술교육대학교)

학술위원장
박상철(대구대학교)
함주연(조선대학교)

프로그램위원장
정남호(경희대학교)
윤상혁(동국대학교)

대학(원)생 아이디어 공모전 위원장
최재원(순천향대학교)
권보람(조선대학교)
김형진(명지대학교)

Agenda

09:00 - 09:30	**등록 및 1차 경품 추첨**				
09:30 - 09:55	**Keynote 01** 스타트업이 미래다: 스타트업 투자와 유니콘 육성 이한수 | 인포뱅크 대표 파트너				
09:55 - 10:20	**Keynote 02** Agentic AI 시대의 신뢰와 거버넌스 – 책임 있는 AI로의 전환 공성배 | 메가존클라우드 부사장				
10:20 - 10:45	**Keynote 03** 무선백도어 해킹의 무한한 진화: AI시대 내부망 보호를 위한 제로 트러스트 전략 이원중 | 지슨 부사장				
10:45 - 11:10	**Keynote 04** AI에게는 난해한 인간의 감정(행복) 서은국 | 연세대학교 심리학과 교수				
11:10 - 12:00	**개회식 및 시상식**				
12:00 - 13:20	**점심**				
	Track 1 라제건홀	**Track 2** 곽정환홀	**Track 3** IBK기업은행홀	**Track 4** 그랜드볼룸A	**Track 5** 그랜드볼룸B
13:20 - 14:00	**Agent AI에 관한 기본적 법률 이슈 검토요건** 최주선 네플라 주식회사 CEO	**AI 시대의 정보 보안과 거버넌스** 최경호 (주)씨커스 대표	**AI 거버넌스와 조직 · 산업 혁신** 좌장: 김남일 연세대학교 교수	**대규모 언어모델과 ESG 경영** 좌장: 이규민 동국대학교 교수	**생성형 AI 및 지능형 플랫폼** 대학(원)생 아이디어 공모전
14:00 - 14:40	**초소형 탈중앙화 고밀도 데이터센터와 온프러미스 AI를 통한 혁신적인 인공지능 생태계 구축** 조정현 주식회사 제타큐브 대표	**토스뱅크 정보보호 거버넌스 사례** 이정하 토스뱅크 CISO	**AI와 신뢰 · 프라이버시** 좌장: 이은곤 수원대학교 교수	**온라인 플랫폼의 위험관리 전략** 좌장: 이중엽 소프트웨어정책연구소 책임연구위원	**핀테크 및 블록체인** 대학(원)생 아이디어 공모전
14:40 - 15:00	**네트워킹**				
15:00 - 15:40	**불확실성 시대에서의 AI 거버넌스와 리스크관리 전략, 새로운 AI 감사 자격 체계 AAIA를 중심으로** 류종기 EY한영 상무	**AI-to-AI 보안의 시대 - 차세대 디지털 신뢰 인프라 제안** 이태훈 (주)매드업 테크리드	**AI와 인간 경험** 좌장: 장재영 한국인터넷진흥원 연구위원	**전기차관련 인식 및 구매 의사결정** 좌장: 고준 전남대학교 교수	**AI 기반 신기술 응용** 대학(원)생 아이디어 공모전
15:40 - 16:20	**AI 시대의 안전한 의료 AI를 위한 책임과 이슈 부제: 고영향 AI 확인 및 사업자 책무 가이드라인을 중심으로** 유석 유니닥스 본부장	**다양한 내부통제에도 불구하고 반복되는 IT 미비점의 구조적 원인 분석: 3개 금융회사를 IT 내부감사 사례를 중심으로** 김재영 딜로이트 안전회계법인 이사	**온라인 리뷰와 eWOM** 좌장: 박하은 경희대학교 교수	**브랜드 스토리와 AI 인플루언서** 좌장: 윤상혁 동국대학교 교수	**AI 비즈니스 모델** 대학(원)생 아이디어 공모전
16:20 - 17:00	**Agentic AI 시대의 서비스 경험 혁신과 신뢰 거버넌스** 김정훈 삼성전자 Principal Professional	**최신 보안 위협 대응을 위해 무엇을 할 것인가?** 박형근 시큐리티플러스커뮤니티 리더	**가짜 리뷰 및 허위 정보 탐지** 좌장: 박재영 국립부경대학교 교수	**디지털 업무환경과 서비스** 좌장: 박상철 대구대학교 교수	**AI와 교육** 대학(원)생 아이디어 공모전
17:00 - 17:30	**시상식, 경품 추첨 및 폐회**				

❖ 본 프로그램은 사전 공지없이 변경될 수 있습니다.
❖ 본 행사는 CPE 7시간이 인정됩니다.
❖ 컨퍼런스에 참석하시는 모든 분들께 점심 식사와 발표자료가 제공됩니다.
❖ 본 행사는 주차비가 지원됩니다.

Keynote Speech 01

스타트업이 미래다: 투자와 유니콘 육성



이한수

인포뱅크 대표 파트너

발표 개요

AI, 반도체, 바이오, 메디컬, 플랫폼 등 스타트업의 최근 투자 동향

이력

이한수 파트너는 카이스트 재학시절 기숙사에서 룸메이트와 싸이월드 창업을 시작으로 전국학생벤처연합회 회장을 지낸 학생 벤처 1세대로서 25년 이상 벤처/스타트업/대기업/투자회사에서 창업/투자유치/전략/마케팅/해외사업/투자집행 등의 경험을 토대로 국내외 주요 스타트업 컴퍼니빌딩 및 투자를 진행하고 있고, 투자한 포트폴리오 기업들의 투자유치/사업전략/마케팅/해외사업 등을 돕고 있고, 중대형 펀드 결성을 위한 LP 업무를 총괄하고 있다. 벤처기업과 스타트업이 대한민국의 미래라고 생각하며 스타트업 생태계에 몸담고 있다.

infobank

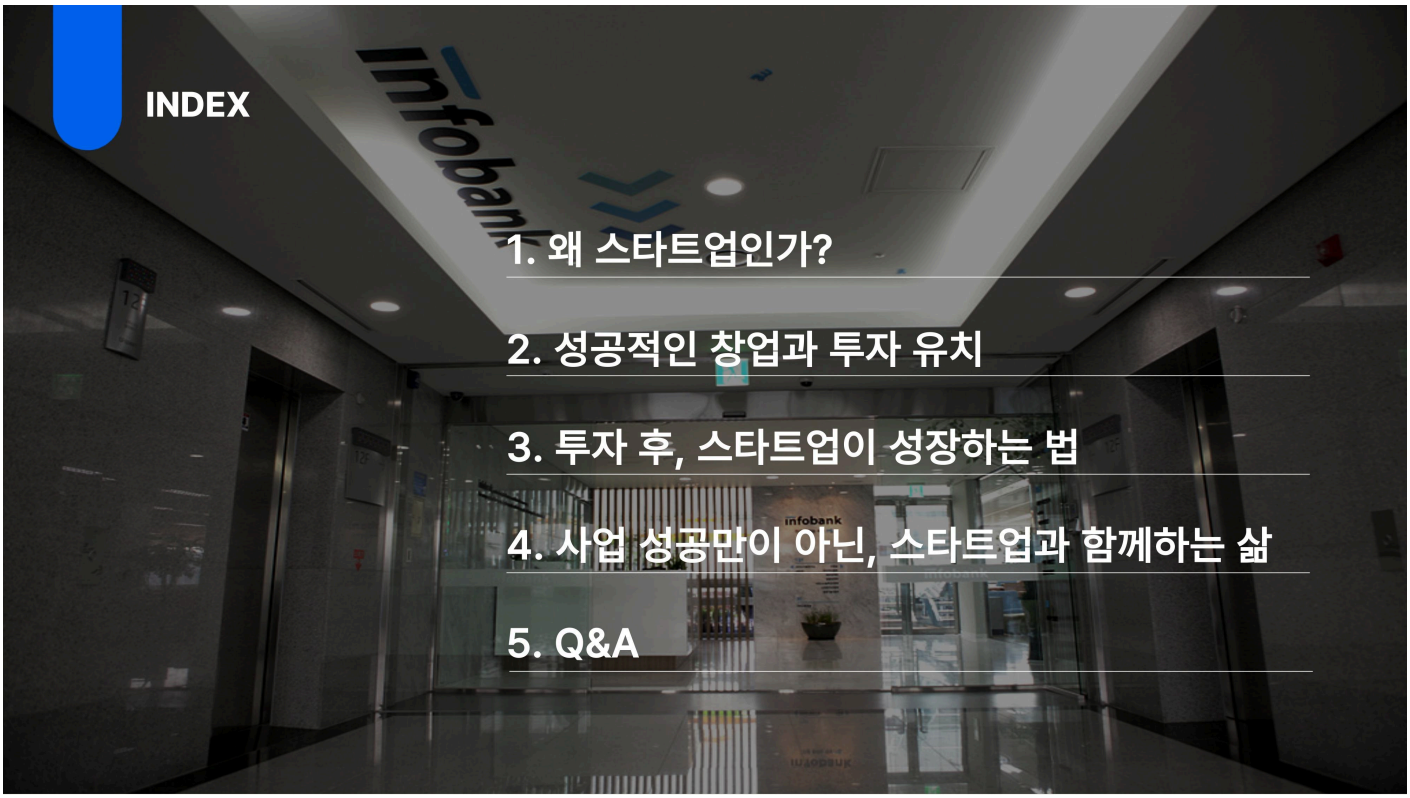
한국지식경영학회

스타트업이 미래다!

스타트업 투자와 유니콘 육성

이한수 대표 파트너

2025. 11



infobank

점 Dot



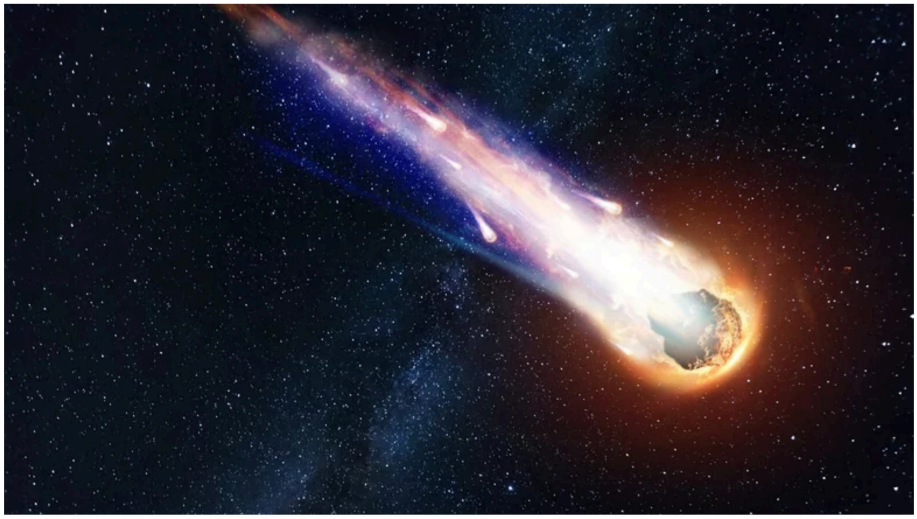
1977.9.5 - 1990.2.14, 보이저 1호, 명왕성, Blue Pale Dot, 창백한 푸른 점, 칼 세이건

2

© INFOBANK. All right reserved.

infobank

Shooting Star



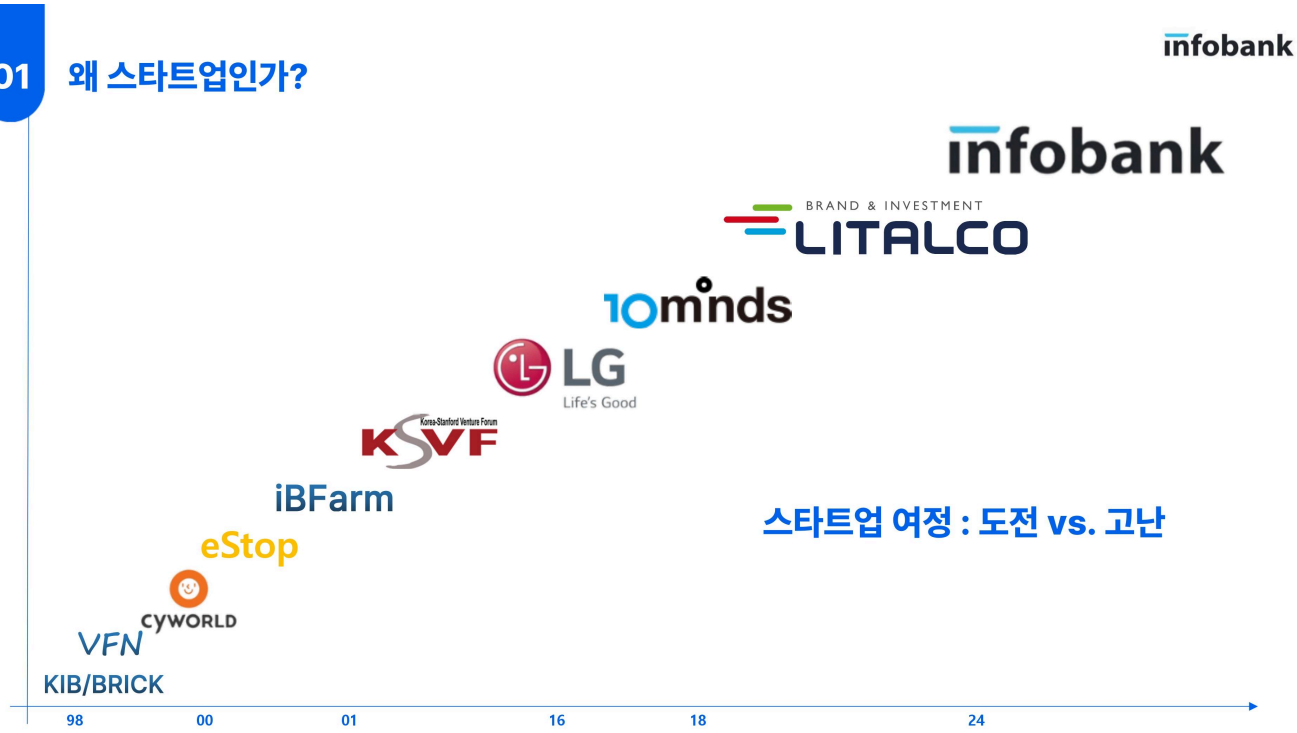
2025.9.10, 춘천 라데나

3

© INFOBANK. All right reserved.

01 왜 스타트업인가?

infobank



4

© INFOBANK. All right reserved.

01 왜 스타트업인가?

infobank

CYWORLD

싸이월드의 탄생 [편집]

1998년에 서울 홍릉의 KAIST 테크노경영대학원에서 이동형, 형용준, 이한수 등 석박사 과정 6명이 EBIZ클럽이라는 창업 동아리를 결성하고 인터넷 관련 공동 창업을 하기로 하였다. 이때 첫 번째 회사인 디지털토네이도를 설립하여 광고 마케팅 서비스인 애드샷을 오픈하였으며 이동형이 대표이사를 맡았다. 한편 형용준은 개인적으로 "peoplesquare.com"이라는 도메인의 사이트를 만들고 일촌 관계의 기능을 도입하였다. 이후 외부 투자자가 피플스퀘어에 관심을 보여 법인 설립을 하기로 한다. 그리하여 1999년에 6명의 회원들 중 3명은 디지털 토네이도에, 3명은 싸이월드에 참여하게 된다. 이한수는 실리콘벨리 투자자 마이클 김을 만나 환타지게임으로 많이 알려진 eLeague를 중심으로 인터넷 커뮤니티 이스탑(eStop)과 글로벌 인큐베이팅사 아이비팜(iBFarm)을 창업하고 LG의 신사업팀장을 거쳐 (주)리탈코(LITALCO)를 창업하였다. 1999년 9월 1

5

© INFOBANK. All right reserved.

01 왜 스타트업인가?

infobank

한경 한경코리아마켓 한경글로벌마켓 집코노미 오피니언 경제 금융 산업 정치 사회 국제 회원전용 전체뉴스 더보기

산업

이스탑, 실리콘밸리서 15억 유치..한국계 텔레비디오사 등서

입력 2000.01.27 00:00 수정 2000.01.27 00:00 가가

한국에 진출한 재미교포 벤처기업이 황규빈 텔레비디오 회장 등 미국 실리콘밸리 투자자 등으로부터 벤처자본을 유치하는 새로운 투자모델이 나왔다.

매일경제 이스탑, 실리콘밸리 자금 15억 유치

기업

이스탑, 실리콘밸리 자금 15억 유치

입력 : 2000-01-27 11:15:17 가

= 엔터테인먼트 콘텐츠 기반 커뮤니티 허브 사이트 운영

[매일경제] 이스탑, 실리콘밸리 자금 15억 유치

<매일경제 이계진 기자> 인터넷 벤처기업 이스탑(대표 김장원 www.eStop.co.kr)은 나스닥 상장사인 텔레비디오의 황규빈 회장을 포함한 실리콘밸리 한국계 투자자들로부터 8억원을 유치하는 등 모두 15억원의 자금을 유치했다고 27일 밝혔다.

황규빈 회장은 최근 인터넷뉴스 C넷에서 인수한 쇼핑몰 비교검색 엔진 마이사이먼(mysimon.com)에투자해 대성공을 거두는 등 엔젤 투자자로서도 왕성한 활동을 하고 있다고 이스탑측은 설명했다.

이스탑은 실리콘밸리 벤처투자 현장에서 실무를 담당해온 마이클 김씨(한국명 김장원)가 KAIST 테크노경영대학원 출신의 이한수 씨와 함께 지난해 9월 설립한 인터넷 업체로 스포츠 영화 연예 미술 등을 포괄하는 본격 커뮤니티 사이트 개발에 주력해왔다.

현재는 인터넷 판타지 농구게임 이리그(eLeague.co.kr)와 2000여명의 연예인 지망생 웹페이지가 있는 이캐스팅(eCasting.co.kr)사이트를 운영중이다.

오는 2월 1일에는 국내 최대 영화 예고편 동영상을 보유한 프리무뷰(PreMoview.co.kr)와 인터넷 화랑 이갤러리(eGallery.co.kr)를추가 오픈해 4개 사이트의 콘텐츠를 기반으로 한 커뮤니티 허브 사이트를 정식 오픈할 예정이다. (02)3453-4131

6

© INFOBANK. All right reserved.

01 왜 스타트업인가?

infobank

매일경제 구독

아이비팜, 한.중 교차 진출 인큐베이팅 사업 실시

입력 2000.11.13. 오전 11:31

공감 댓글 가가

= 벤처 25사 中진출 인큐베이팅 =

<정현권> 한국과 중국 벤처기업의 양국간 교차진출을 매개하는 글로벌 인큐베이팅 업체가 탄생했다. 아이비팜글로벌(www.ibfarm대표이종혁 이한수)은 최근 25개 한국 벤처기업의중국 진출과 5개 중국벤처기업의 한국 진출을 위한 인큐베이팅 사업에 들어갔다.

중국 진출을 추진하는 업체로는 화훼 B2B(기업간 전자상거래) 전문기업인 조이인박스(www.joyinbox)와 건축 및 인테리어 설계용 3차원가상현실 소프트웨어 업체인 한국가상현실(www.kovi.com)등 25사다.

한국에 진출하려는 중국 벤처기업은 무선 네트워크업체인 상하이 신국정보통신과 모바일 문자압축인식기술 업체인 비지텍, 중국 최대 아동 포털사이트인 중국아동망 등 5사다. 특히 중국아동망과 한국 키드앤키드의 콘텐츠교류 및 아동게임 솔루션의 공급을 결정, 양국간 아동문화사업을 적극 추진키로 했다.

이 회사는 양국간 벤처인큐베이팅 사업을 활성화 하기 위해 `아이비팜 차이나`라는 법인을 세우고 상하이와 베이징에 3개 인큐베이팅센터를 마련, 인큐베이팅사업에 들어갔다.

특히 아이비팜 차이나는 최근 e-삼성차이나와 이 사업과 관련해 제휴를 맺고 정보통신기술 및 인터넷 사업의 네트워크를 공유하기로 했다.

매일경제 구독

아이비팜, 中 진출 전담지원 벤처

입력 2001.06.07. 오전 11:59

공감 댓글 가가

<정현권>벤처 인큐베이팅업체를 통한 벤처기업의 중국 진출이 활성화되고 있다.중국의 법적.제도적 배경에 대한 경험과 지식이 미비해 전문 인큐베이팅업체를 통한 대중국 진출이 벤처업계의 새로운 대안으로 떠오르고 있는 것.

글로벌 인큐베이팅업체인 아이비팜(www.ibfarm.com 대표 이종혁.이한수)는 최근 중소기업청에서 '벤처기업 해외진출을 위한 중국 사업수행기관'으로 선정됐다. 이에 따라 이 회사는 앞으로 중소기업청이 지원하는 100여 해외진출 희망 벤처기업에 대해 중국 진출 전담 지원기관 역할을 하게 된다.

이를 위해 이 회사는 1년여에 걸쳐 현지조사와 네트워크를 구축하면서 지난해 한국과 중국 상하이에 각각 현지법인을 설립했다. 중국법인에는 현지 전문 인력들을 채용해 시장기회분석 및 조사, 솔루션 현지화, 판매및 협력 파트너 발굴, 현지법인 설립 등의 업무를 지원하고 있다.

7

© INFOBANK. All right reserved.

01 왜 스타트업인가?

infobank



© INFOBANK. All right reserved.

01 왜 스타트업인가?

infobank

천만 관객 관람 영화



조연 오달수

천만 요정

조연 인포뱅크

유니콘 요정

© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank

인포뱅크의 투자 사업 현황

- **350 개** — 전체 투자 포트폴리오사
- **950억원** — 누적 투자 금액 (고유 계정 15%, 펀드 계정 85%)
- **235개** — 팁스(TIPS) 선정 기업 (2016 ~ 2025)
- **60여개** — 연간 투자 기업 수 (매월 5~7개사)
- **95% 이상** — 투자 기업 생존율
- **5.5조원 이상** — 투자 포트폴리오사의 총 기업 가치
- **71%/3천억원↑** — 후속투자 성공율[1)]
- **5배 이상** — 후속투자유치 기업의 평균 기업가치 상승율[1)]
- **2,933건** — 파트너사 IP지원

*1) 투자 후 1년 미만 제외

© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank

- **파트너사 누적 후속 투자유치 금액 3,000억 원 초과, '23년 쏘카 외 2개사, '25년 TXR로보틱스 IPO 입성**

투자 성공 사례

IPO 성공 사례(6개사)

com2us 컴투스 · SOCAR · ASTA · Qualitas · crowdworks · TXR ROBOTICS

IPO 예정('25~'27, 8개사)

VINSSEN · buzzvil · RIDE · waycen · colosseum · AP MDS Co., Ltd. · POSICUBE

M&A 성공 사례(6개사)

buzzvil · neptune · SJM · SOCAR · 모두의주차장 · JUST CO · ALB · VAXCELL

아기유니콘(9개사)

QS TAG · colosseum · CABANG · waycen · BITE ME · bejewel · XINAPSE · Wayne Hills Bryant A.I · HIGHER

본계정 투자 사례

"182억원 기 회수(2.5배수) / 회수 예정 포함시 407억원(5.6배수)"

본계정 시즌1 (14~19년) 투자 → 73억원 투자 / 67개사 → 182억원 회수 완료(11개사) + 225억원 회수 예정(40개사)

인포뱅크 개인투자조합 1호 사례

"투자수익 Multiple 5배수(IRR 32.7%)"

19년 8월 설립 결성총액50억원 (6년 경과) → 42.5억원 투자 / 19개사 → 기업가치 6.7배/ 투자수익 5.0배 IRR(32.7%)

© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

Overview "Global Logistics DX Partner, 콜로세움"

colosseum
Logistics Simplified, Success Delivered

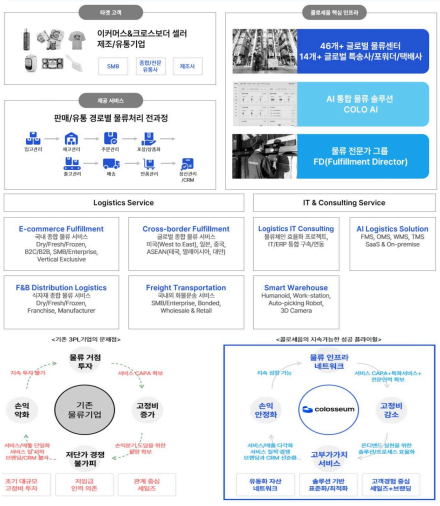
주식회사 콜로세움코퍼레이션

infobank

대표이사	박진수
설립일	2019.05.22
직전 기업가치/누적투자금액	Pre 850억 원 / 207억원

제품 · 서비스 소개

콜로세움은 고도화된 AI 솔루션을 통한 물류효율화를 이루며 이에 더해 전국의 네트워크화된 물류센터를 활용하여 이커머스 상품의 보관, 주문처리, 포장, 배송, 반품 처리까지 원스톱으로 지원해주는 풀필먼트 서비스를 제공함



목표 시장 규모 및 전망

- 유통 소매시장 연 평균 6.4% 성장
- 글로벌 3자물류 시장규모 : $0.97 Trillion – 1.14% of total global GDP in 2030



투자 포인트 / 경쟁력

커스텀 역량과 서비스 안정성 (온디맨드 풀필먼트)

- 복잡다양한 물류니즈의 유연한 해결 : COLO 기반의 가시성 + 특화 물류센터 네트워크 + FD 밀착케어 + 최대 30% 저렴한 비용
- 경쟁사 불만족 문의고객 비중: 30%이상 유지 ('23~24년)
- 고객사 서비스 만족도: 80%이상 유지 ('22~23년)
- 고난이도 물성 해결역량 강화
  : 부가서비스 평균 이용갯수 1.9개('21년) → 2.3개('23년)

매출유형 다각화와 성장속도 (글로벌 확장에 유연한 비즈니스 모델)

- 연 평균매출 성장률 191%('20~24년), 창업후 6년간 누적매출 600억원 달성을 달성하며 빠르게 성장 중
- SME, Enterprise와 Cross border, Consulting, 화물주선, 식자재, 물류컨설팅, SaaS솔루션 등 매출의 구성비 다변화
- 미국 LA 풀필먼트 센터 구축 및 COLO Global 운용
- COLO SaaS를 통한 솔루션 매출의 확대(국내외 포함)

수익성 확보 및 개선 (목표 영업이익률 8%, 기존 물류기업 평균 2%)

- '24년 4분기 월 BEP 달성, '25년 당기순이익 예상

매출 및 이익 목표

(단위: 억원)

년도	2023	2024	2025(E)	2026(E)	2027(E)
매출액	118	395	2,490	4,839	9,405
영업이익	△17	△26	91	428	1,144

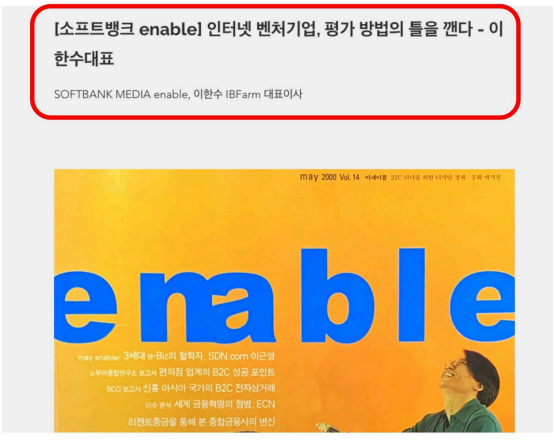
대표자 및 주요 인력 소개

이름	역할	소개
박진수	CEO	• 사업기획 / 경영 총괄 (경력 17년) • KTF / KT 개인마케팅전략 담당 • 대학내일 경영진 • 서울대학교 경영학/소비자학 학사
유인형	CPO	• 프로덕트 기획/설계 총괄 (경력 16년) • KTF / KT 신규서비스 기획 담당/ 멋진경험 대표이사 • 서강대학교 경영학 학사
윤태식	CSO	• 셀러 유치 / 창고 영업 총괄 (경력 16년) • 수입무역 / 제휴 및 유통 전문 / 멋진경험 이사 • 중앙대학교 경영대학원 유통산업학 석사
송유철	CFO	• 경영관리총괄 (경력 18년) • Goldman Sachs, Morgan Stanley Finance • EY 한영회계법인 감사본부 • 부산대학교 경영학 학사

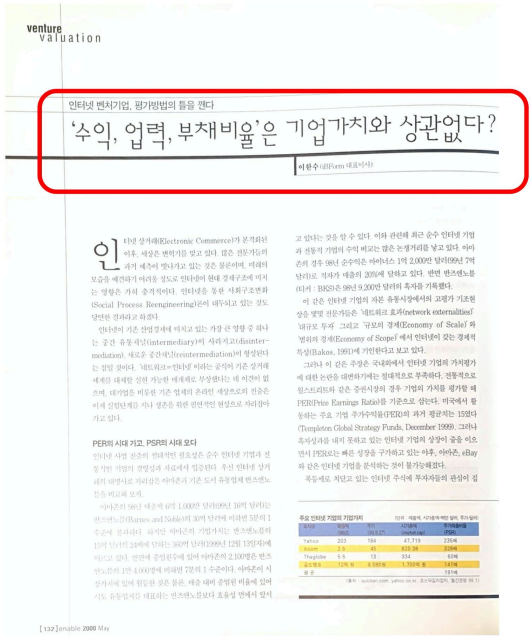
© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank


[소프트뱅크 enable] 인터넷 벤처기업, 평가 방법의 틀을 깬다 - 이한수대표

SOFTBANK MEDIA enable, 이한수 IBFarm 대표이사


venture valuation

인터넷 벤처기업, 평가방법의 틀을 깬다

'수익, 업력, 부채비율'은 기업가치와 상관없다?

이한수(IBFarm 대표이사)

© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

AP MDS Co., Ltd.
주식회사 에이피엠디에스

infobank

대표이사	김종정
설립일	2015.08.24
직전 기업가치 / 누적투자금액	Post 307억 원 / 80억 원

Overview "The only total anti-infective solution provider" from material to product "

제품 · 서비스 소개

에이피엠디에스는 세균감염의 주요 원인인 생물막 형성을 원천적으로 차단하는 Bi-Fi Free 소재 합성기술과 이를 응용한 소모성의 의료기기 및 항생물막 제품, 소재에 이르기까지 "감염방지 Total Solution"을 제공하는 유일한 기업입니다

1. 비뇨기용 항생물막 카테터
- 2WAY FBC(6Fr~26Fr) / 3WAY FBC(14Fr~26Fr)
- 협착 방지 및 통증 감소 4-way Membrane Catheter

2. 감염 예방 항생 실리콘 소재 및 의료 소모품
- Medical Silicone Bandage / TOMCAT Catheter / 영유아용 제품 등

3. 특징

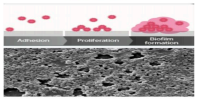
① 세계 최초 Silicone과 항생력을 가진 ZnO Polymer를 특허기술로 합성한 항생 소재

② 생체 내 활성전위를 이용한 정전 효과(Galvanic effect)로 원인균을 비 활동성으로 변화시켜 생물막 형성을 차단하여 원천적으로 증식 차단

③ 합성된 소재의 Wet Slip 특성으로 생물막 형성에 의한 협착 및 통증을 현저히 감소시킴

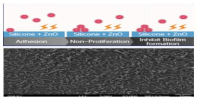
④ 비뇨기용 범용 튜브카테터의 경우 항생 기능이 없어 항생제나 항생물질을 코팅하는 방법이 있으나 장기적 요도에 유치 시 항생제 내성, 요도 협착 및 코팅 금속의 박리에 의한 부작용 발생

일반 실리콘



군집을 통한 생물학 형성

아폴론 항생 실리콘



생물학 형성 억제

목표 시장 규모 및 전망

(시장 규모)
- '23년 Catheter 세계 시장 규모는 약 380억 달러 규모이며, 비뇨기과용 카테터의 비중은 37%로 약 140억 달러 규모임
- 실리콘 기반 소모성 의료기기 시장 규모는 연간 500억불 이상이며, 매년 7% 이상 성장하고 있음

(시장 전망)
- 실리콘 기반 소모성 의료기기 제품 중 50% 이상이 항생물막 기능 요구 **(주요 시장인 유럽, 미국 진출을 위해 CE, FDA-510k 인증 완료)**
- '22년 5월 SCI급 논문 발표 / '23년 10월 신제품(통증감소약물분사 카테터) 개발 완료 및 통증감소 입증 논문 발표 / 치료 재료 수가의 비보험 (비급여) 진행 완료 예정(25년 3월)
- 글로벌 기업들과의 다양한 협업(OEM 생산, 인프라 및 네트워크를 활용한 마케팅 전략, 반제품 등 수출, Collaboration model 개발 등)을 통한 매출 확대

투자 포인트 / 경쟁력

- **소재 자체에서 항생 기능을 발현하는 원천기술을 보유**하여 경쟁사에서 모방이 어려운 소재 기술로써 **기술적 장벽이 높음**
  (국내특허 등재 6건 및 미국과 중국 PCT 2건 출원 및 등록)
- 일반 실리콘 제품 대비 **2배 이상의 항생 기능 보유**, 소재 특성 상 타사의 제품과 비교하여 부작용 리스크(항생제 내성, 중금속 박리 등) 문제 해소
- **KFDA, CE, 미국FDA 인증 완료**
  - 국내 뿐 아니라 유럽, 동남아시아, 미국 등 진출이 가능하며 미연방 조달 업체(SAM) 등록 및 공급 시작
- 고유의 항생 기능을 가진 실리콘으로 소모성 의료 기기 시장 뿐 아니라 **상업용 실리콘 시장 확장 가능**
- 감염을 원천적으로 차단하는 소재 및 제품 기술로 미래 가치 높음
- **미 연방 조달 시장 진출**
  - **Surgical 제품으로는 미국기업 외에 제조기업 VA(미 보훈처) 최초 승인**
  - **VISN-5(워싱턴) 구역 확정 및 VISN-21(캘리포니아) 구역 진출 예정**

매출 및 이익 목표

(단위 : 억원)

년도	2023	2024	2025(E)	2026(E)	2027(E)
매출액	5.5	10.5	112	228	250
영업이익	△12	△2	40	78	80

대표자 및 주요 인력 소개

이름	역할	소개
김종정	CEO	• 연세대 의학공학 석사 • 씨유메디칼(코스닥 상장사) 연구센터장 역임
전용한	전무	• 씨유메디칼 RA 경력 12년 • 의료기기 제품 인허가
우건욱	상무	• 씨테크 반도체 차폐 공정 R&D 경력 15년 이상 • 소재합성 분산공정 및 품질 총괄
박명재	이사	• 씨테크 반도체 제조 공정 경력 15년 • 제조 공정 총괄

© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank

보유 펀드 현황

총 24개 1,300억원 규모 블라인드 및 프로젝트 펀드 운용(모태펀드 8, 성장금융 1, 비모태펀드 15개)

모태펀드 & 성장금융 (FoF)

개인투자조합 (4)

- 인포뱅크 개인투자조합 1호
- 인포뱅크 개인투자조합 2호
- 인포뱅크 개인투자조합 3호
- 인포뱅크-대구창조경제혁신센터 개인투자조합 1호

벤처투자조합 (5)

- 인포뱅크 핀테크 혁신펀드 1호
- 인포뱅크 창업초기 혁신펀드 1호
- IB-BDC 공공기술사업화 혁신펀드 1호
- 인포뱅크 미래환경 혁신기술 펀드
- 아 IB 초격차 펀

프로젝트 및 블라인드 펀드 (14)

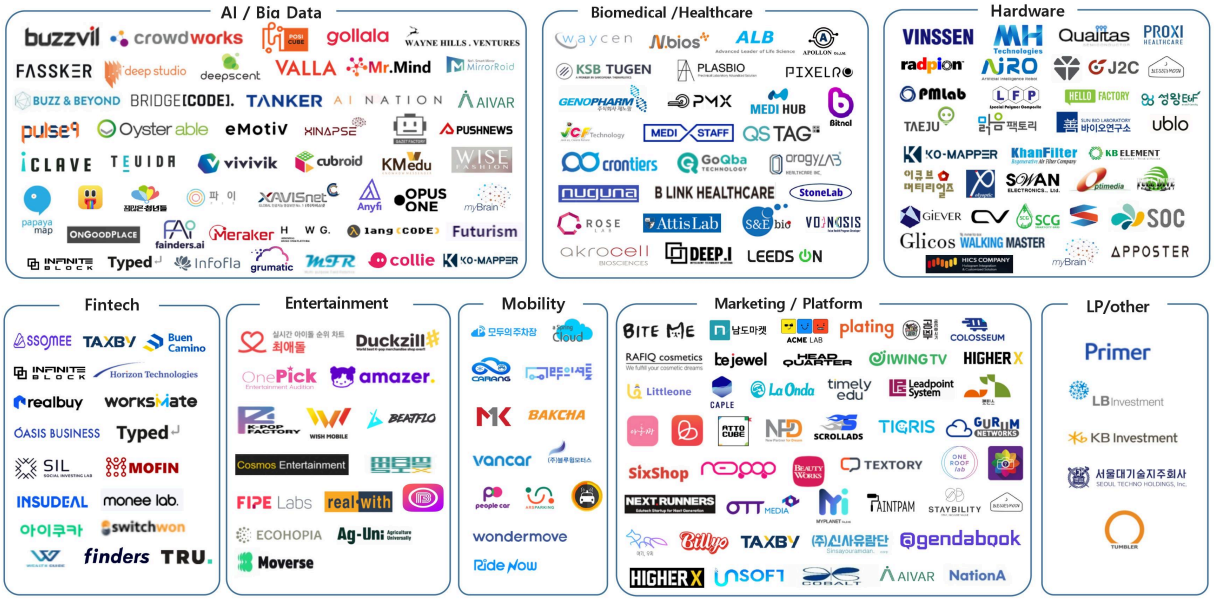
- 바이오 — 바이오 개인투자조합 1호
- ESG — ESG 개인투자조합
- 메타버스 — 메타버스 개인투자조합
- iAccel I~IV — iAccel 개인투자조합 1호
- SPAC — SPAC 개인투자조합 1호
- 벤처이노베이션 — 벤처이노베이션 개투투자조합
- 삼성화재 — 삼성화재-인포뱅크 창업초기펀드
- 서울바이오 — 메디컬 개인투자조합 1호
- 99도씨 — 99도씨 투자조합 1호
- 삼성화재 — 밸류업 벤처투자조합
- iAccel V — iAccel 개인투자조합 5호
- 엘지생건 — 엘지생건 혁신조합 1호
- 프리IPO — 인포뱅크파트너스 1호

© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank

현재 AI, 빅데이터, 핀테크, 바이오/헬스케어, 하드웨어, 모빌리티 등 전분야에 걸친 투자 포트폴리오사 보유



© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank

'22년, 23년 팁스 선정 1위의 10년차 팁스운영사, '팁스, 딥테크 팁스 프로그램 등' 팁스 연계형 투자와 결합하여

투자기업에게 충분한 자금을 공급함으로써, Death Valley 도약 지원 및 성공 가능성을 높이는데 집중 가능 !

팁스 지원을 통한 극초기기업의 Runway 확보

- 운영사는 2015년부터 본격적인 팁스 운영사로 활동하면서 총 227개의 팁스 선정기업을 선정하였으며, 투자금액 약 500억 원 대비 R&D 정부지원금 약 920억 원 매칭에 성공함

<운용사 연도별 팁스 배출 기업 현황>

구분	16년	17년	18년	19년	20년	21년	22년	23년	24년	합계
기업수	6	6	16	11	17	26	34	45	60	227

- 2022년에는 31개사, 2023년에는 42개사(딥테크팁스 12개사 포함)의 팁스 기업 배출을 통해 전체 104개 팁스 운영사 중 1위를 기록함



- 제안펀드를 통해 3억 원 이상 투자 시 팁스 및 딥테크팁스 프로그램을 통해 최대 17억 원의 자금 매칭이 가능하므로, 창업초기 기업들의 Runway를 길게 늘려줌으로써 Death Valley 도약을 지원함
- 또한 자금 확보를 통해 브릿지 투자 효과를 조기에 달성하고 시리즈 A 투자 시기를 단축시킬 수 있음



운용사의 주요 팁스 선정기업 기업가치 상승분

- 주요 15개사 평균 40배수 성장 -

No	Company	투자 시기	Initial Valuation(a)	Current Valuation(b)	Multiple (b)/(a)	Comment
1	퀄리타스반도체	2018-11-15	20	5,814	290.7	시가총액
2	크라우드웍스	2018-04-02	25	3,435	137.4	시가총액
3	비주얼	2017-09-27	11	881	80.1	3자거래가
4	에이피엠디에스	2016-03-25	5	286	57.2	3자 거래가
5	스프링클라우드	2017-10-31	10	416	41.6	3자 거래가
6	딥스튜디오	2018-06-27	17.5	497	28.4	3자 거래가
7	플랫팜	2017-06-29	7	180	25.7	3자 거래가
8	메디허브	2018-06-15	10	228	22.8	3자 거래가
9	포지큐브	2018-05-15	32	690	21.6	3자 거래가
10	웨이센	2020-02-21	38	919	24.2	3자 거래가
11	빈센	2019-10-18	48	870	18.1	3자 거래가
12	웨인힐스브라이언트에이아	2020-08-26	43	521	12.1	3자 거래가
13	에프앤에스홀딩스	2019-10-18	49	710	14.5	3자 거래가
14	콜로세움코퍼레이션	2020-06-08	36	850	23.6	3자 거래가
15	큐에스택	2019-07-08	26	288	11.1	3자 거래가
합계			377.5	16,585	44.0	-

투자 및 팁스 선정 후 70~80% 기업이 1~2년 내에 약 3배수 성장했으며,

추가 1~2년 경과 후 시리즈 B 단계로 약 10배수 이상 성장

© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank

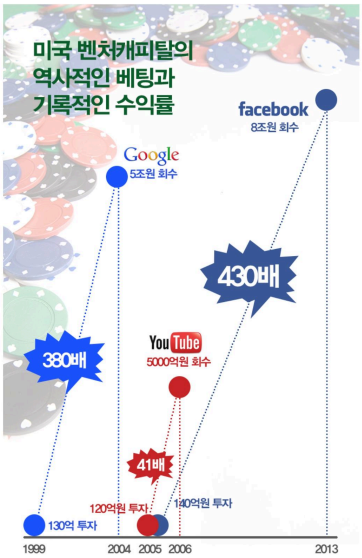
전체 포트폴리오사의 50% 이상이 팁스/딥테크 팁스 선정 기업, ▶ 총 235개사 (딥테크 팁스 선정팀 21개, 운영사 중 최대 보유)



© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank





If you don't find a way to make money while you sleep, you will work until you die.

— Warren Buffett —

창업과 투자 원칙

빨리 가려면 **혼자** 가고
오래 가려면 **함께** 가라!

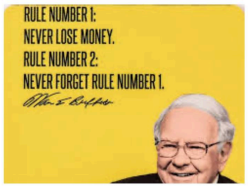
투자는 연애처럼
창업은 결혼처럼!
w/ 경험, 동반자

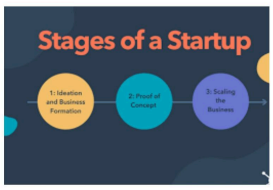
사업은 창업자가
투자는 투자자가
초기투자는 인포뱅크!

© INFOBANK. All right reserved.

02 성공적인 창업과 투자 유치

infobank

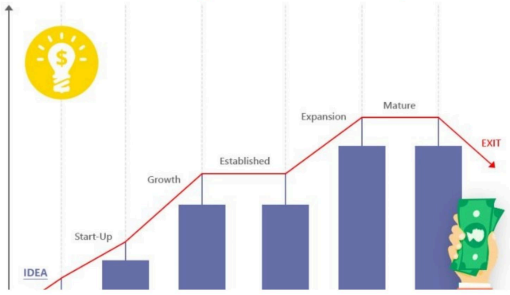




사업 원칙

1. ALWAYS MAKE MONEY!
2. NEVER FORGET RULE NO.1

The 7 Stages of the Startup



투자 원칙

I. **대표**와 **창업팀** 역량 (fit)
II. **시장** Size / **성장성** (숫자, exponential)
III. BM 사업화/**개발** 가능성 (1st mover)
IV. 팁스 (TIPS) 가능성 : 딥테크/스케일업팁스
V. Follow-on **투자** / **EXIT** 가능성

© INFOBANK. All right reserved.

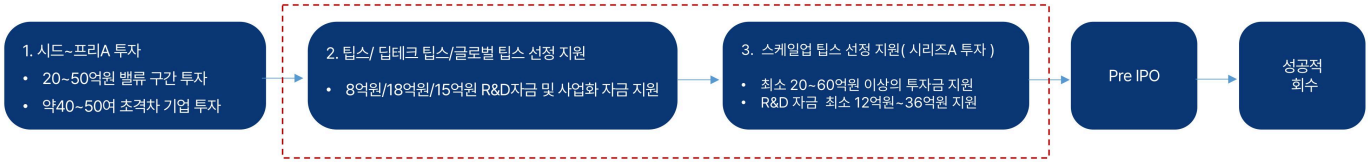
03 투자 후, 스타트업이 성장하는 법

infobank

" 프리 팁스 운영사 ~ 팁스/딥테크 팁스 운영사 ~ 스케일업 팁스 운영사로 활동 중, 팁스 전주기 밸류체인 완성"

팁스 - 스케일업 팁스 연계 지원

- **스케일업 팁스 컨서시엄(인포뱅크, 우리벤처파트너스, 교보증권)에서 인포뱅크 팁스/딥테크 팁스 선정팀에 최소 10억원 이상 투자시, 한국벤처투자 매칭펀드로 최소 10억 이상 매칭 투자 지원 및 R&D 지원금 12억원 제공**
- **DCP(딥테크 분야) 기업의 경우, 컨소시엄에서 20억 이상 투자시, 한국벤처투자 매칭투자 40억원+R&D 자금 36억 지원**



팁스~스케일업 팁스 지원을 통해 투자기업에게 (최소 20~60억원 이상의 매칭투자 자금 + 최소 20~44억원 R&D 자금) 지원 가능~
시리즈B단계 투자 없이 Pre IPO 가능성 제고 + 100배 이상의 멀티플 수익 창출 가능!

© INFOBANK. All right reserved.

03 투자 후, 스타트업이 성장하는 법

infobank

짧은 시간 안에 폭발적으로 성장한 투자 포트폴리오 Value-Up 사례

투자업체	Value-up 사례
크라우드웍스 (18년4월 투자)	▪ '영상 인식을 위한 인공지능 학습용 데이터셋 마켓플레이스 개발' 과제로 팁스(TIPS) 선정 지원 (2018년 06월 선정) ▪ 국내 특허 출원 100건, 해외 특허 출원 14건, 국제특허(PCT) 14 건 등 데이터 라벨링 관련 기술로 국내 최다 특허 보유 ▪ 투자 시점 기업가치 25억 원, 상장 후 최고가 기업가치 3,435억 원 **(137배수 이상 성장)**
포지큐브 (18년5월 투자)	▪ '클라우드 기반 대화형 인공지능 예약 플랫폼 개발'과제로 팁스(TIPS) 추천 및 선정 지원(2018년 05월 선정) ▪ 포지큐브 후속 투자 유치 지원 (캡스톤파트너스, 원베스트벤처투자, HB인베스트먼트 추천 및 투자 유치 성공) ▪ 투자 시점 기업가치 32억 원, 현재 기업가치 690억 원 **(22배수 이상 성장)**
스프링클라우드 (17년10월 투자)	▪ 기업공개(IPO) 준비 중(상장 주관사: 신한금융투자) ▪ '자율주행 레벨 3용(운전자상태/전장주행인지) 실시간 모니터링 듀얼 카메라 시스템 개발(제품명:SpringGo Model-1)'과제로 팁스(TIPS) 선정 지원(2017년 09월 선정) ▪ IP 컨설팅 지원 (특허 24건, 상표 25건 출원 지원) - 자율주행 차량의 카메라 시스템, 충돌 예방, 변경 장치, 탑승자 모니터링 관련 기술 ▪ 투자 시점 기업가치 10억 원, 현재 기업가치 416억 원 **(42배수 이상 성장)**
엠에이치기술개발 (17년7월 투자)	▪ 인수합병(M&A) 진행 중(상장사 ㈜에스제이엠으로부터 인수합병 추진 중) 전기차 부품 사업 협력 후 지분 확대 및 인수합병(M&A) 검토, 아주IB 250억 밸류 30억 투자 유치 진행 중 ▪ 데쓰밸리에 뼈저 전직원 3인(대표1, 대표 가족2) 만 남은 상태에서 회생 가능한 최소 투자금액 추가 투자 지원을 통해 성공적인 회생 및 후속투자 유치 ▪ 투자시점 기업가치 15억 원, 현재 기업가치 250억 원 **(17배수 이상 성장)**
퀄리타스반도체 (18년11월 투자)	▪ 2023년 10월27일 코스닥 코스닥 상장 (상장 주관사: 한국증권) ▪ '단거리 광통신을 위한 초저전력 광통신 반도체 집적회로 개발' 과제로 팁스(TIPS) 선정 지원 (2018년 10월 선정) ▪ IP 컨설팅 지원 (특허 27건, 상표 6건 출원 지원) – 적응형 등화장치, AOC 장치, 리피터 장치, PAM 신호 처리 회로 관련 기술 ▪ 투자 시점 기업가치 20억 원, 상장 후 최고가 기업가치 5,814억원 이상 **(290배수 이상 성장)**
웨인힐스브라이언트 에이아이 (20년8월투자)	▪ 기업공개(IPO) 준비 중(상장 주관사: 신영증권), 성장성 특례 상장으로 2022년 상반기 코스닥 상장을 위한 본격 준비 ▪ '인공지능 음성인식을 활용한 전문상업영상 자동제작 서비스' 과제로 TIPS(팁스) 추천 및 선정(2020년 05월 선정) ▪ 투자 시점 기업가치 43억 원, 현재 기업가치 521억 원 **(12배수 이상 성장)**

투자업체	Value-up 사례
클로세움 (20년6월 투자)	▪ 'AI 기반 물류 유통 솔루션' 과제로 팁스(TIPS) 선정 지원(2020년 08월) ▪ IP포트폴리오 구축 및 후속 투자유치 연결 지원 ▪ 투자 시점 기업가치 36억 원, 현재 기업가치 850억 원 **(24배수 이상 성장)**
에프엔에스홀딩스 (19년10월 투자)	▪ 'AI 기술 기반 글로벌 명품 패션 플랫폼' 과제로 팁스(TIPS) 선정 지원(2019년 11월), IP 컨설팅을 통해서 특허 22건 출원 지원 ▪ 투자 시점 기업가치 49억 원, 현재 기업가치 710억 원 **(14배수 이상 성장)**
웨이센 (20년2월 투자)	▪ '정밀의료를 위한 인공지능 기반의 소화관 내시경 실시간 영상 분석 솔루션 개발' 과제로 팁스(TIPS) 선정 지원(2020년 03월 선정) ▪ IP 컨설팅을 통해서 특허 20건 출원 지원 ▪ 투자 시점 기업가치 38억 원, 현재 기업가치 919억 원 **(24배수 이상 성장)**
바잇미 (20년1월 투자)	▪ 취향이 있는 반려생활을 만들어주는 펫 브랜드 이커머스 플랫폼 운영 ▪ 후속투자 유치 지원, 누적 93억원, 미국 법인 설립 등 적극적인 온/오프라인 판매 지원 ▪ 투자 시점 기업가치 23억 원, 현재 기업가치 462억원 **(20배수 이상)**
에이피엠디에스스 (16년3월 투자)	▪ '항균기능이 구비된 카테터 개발' 과제로 팁스(TIPS) 선정 지원(2016년 03월 선정) ▪ IP 컨설팅을 통해서 특허 12건 출원 지원 ▪ 투자 시점 기업가치 5억 원, 현재 기업가치 286억 원 **(57배수 이상 성장)**
비주얼 (17년9월 투자)	▪ 투자연계형 R&D사업 수주 지원(3억원 정부과제 수주), 시리즈 A단계 후속투자 유치 지원, 특허 아이템 발굴 지원(특허 18건) ▪ 투자 시점 기업가치 11억 원, 현재 기업가치 880억 원 **(80배수 이상 성장)**
메디허브 (18년6월 투자)	▪ 시리즈A단계 후속투자 유치 지원(IR덱 코칭 등), 제품 판매 홍보/마케팅 지원, 특허 아이템 발굴 지원 ▪ 투자 시점 기업가치 10억 원, 현재 기업가치 228억 원 **(23배수 이상 성장)**
빈센 (19년10월 투자)	▪ '400마력(300kW) 이상 추진력을 가진 미래형 친환경 전기추진보트 개발' 과제로 TIPS(팁스) 추천 및 선정(2019년 12월 선정) ▪ 후속투자 집중 지원을 통해 투자 후 11개월만에 후속투자 유치 성공, 팔로우 온 투자 확정을 통한 후속투자 연계 지원 ▪ 글로벌 진출을 위한 협력체계 구축 및 지원(싱가로프 해외 법인 설립 준비 중) ▪ 투자시점 기업가치 48억 원, 현재시점 기업가치 870억 원 **(18배수 이상 성장)**
플랫팜 (17년6월 투자)	▪ 시리즈A단계, B단계 후속투자 유치 지원(IR덱 코칭, 약30여 투자자 연결 지원 등) ▪ 특허 아이템 발굴 지원(특허 69건) ▪ 투자 시점 기업가치 7억 원, 현재 기업가치 180억 원 **(26배수 이상 성장)**

© INFOBANK. All right reserved.

03 투자 후, 스타트업이 성장하는 법

infobank

총 1조8천6백억원 AUM , 329개 기업 IPO, 4천억원 이상 투자재원 확보, 총 140여건 R&D 기술사업화&정부지원사업, 팁스 1위 운영사, 글로벌 파트너십 보유

투자기관

01 우리벤처파트너스

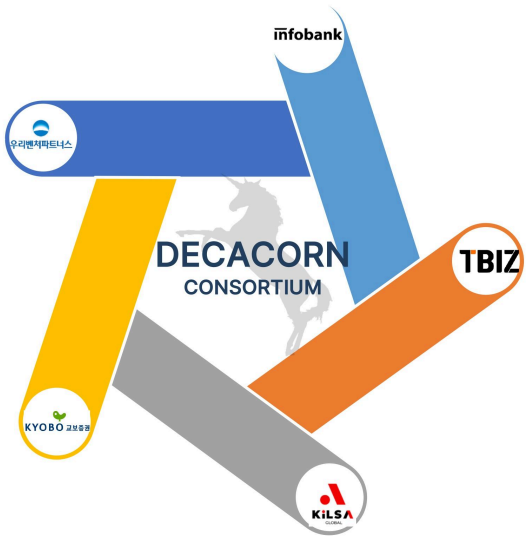
역할
- AUM : 1조 2,971억 원 규모
- Deep-Tech,ICT, Bio 등 특화 투자

강점
- 320개 기업의 IPO 달성 경험
- 글로벌 진출 연계 지원

02 교보증권

역할
- AUM : 4,097억 원 규모
- 초격차 분야 및 지역 특화 투자

강점
- 연평균 25% 성장률의 투자 성과
- CVC기반 금융 네트워크 활용



연구기관

03 인포뱅크

역할
- 창업초기투자및스케일업 프로그램
- 딥테크팁스/팁스 연계 지원

강점
- 팁스 208개(딥테크18개)기업보유
- 약 7백억 규모 투자 실적
- ICT 기술 자문 및 R&D 지원
- 글로벌 진출 연계 지원

04 티비즈

역할
- 기술사업화, R&D 스케일업 지원
- 지역 산업과 협력 네트워크 활용

강점
- R&D 기반 전주기 기술사업화 스케일업 프로그램 운영

05 킬사 글로벌

역할
- 글로벌 진출 현지화 지원

강점
- 3년 10개 기업의 글로벌 IPO 달성

© INFOBANK. All right reserved.

03 투자 후, 스타트업이 성장하는 법

infobank

해외 AC/VC 및 대기업/유니콘 기업의 국내 스타트업에 대한 관심을 제고하고, 전주기 협력 네트워크를 강화함

글로벌 파트너

[싱가폴]



[Origgin MOU]



[Pavilion]

[태국]



[EAC]

[미국]



[Stella]

신규 글로벌 파트너사 발굴 및 구축 현황

- 글로벌 네트워크를 구축 : 분기 1회 이상 글로벌 VC 대상 IR 진행
- 해외 AC/VC 등과의 본격적인 협력을 통해 **2년 이내 글로벌 공동 펀드 1개 이상 결성을 목표**로 함
- 해외 VC/AC의 국내 스타트업에 대한 관심을 제고함으로써 기업의 성장 동력 발굴을 돕고 보육 기업의 해외 진출을 적극 지원할 예정임

구분	파트너사명	국가	세부 사항
구축	Creww	일본	• 일본 현지 및 APAC 권역 CVC 네트워크를 기반으로 한 POC 협력 지원 중이며, 공동 펀드 조성 등을 위한 MOU 체결 진행(24.09)
	EAC	태국	• 태국 투자사 및 스타트업 AC와 협업 통해 포폴사의 비즈니스 협력 지원 중이며, 공동 펀드 조성 등을 위한 MOU 체결 진행(25.01)
	Origgin Ventures	싱가포르	• 팔로온 투자나 공동 투자 등에 있어서 추후 협력 진행 예정이며, 공동 펀드 조성 등을 위한 MOU 체결 진행(24.11)
발굴	LG Electronics Japan Inc.	일본	• 현재 일본 내 연구·개발(R&D) 거점의 인재 확보에 나서며 '초격차' 기술 확보에 총력을 기울이고 있으며 국내외 스타트업 등과도 적극적인 협업을 희망하며 사업화 등에 있어 연계 목표
	Mitsui & Co.	일본	• 전세계 약 66개국에서 자원·에너지 분야부터 미디어 분야까지 산업 전 분야에서의 인프라를 바탕으로 투자와 경영을 활발히 하고 있음 • 각 부서마다 POC 등을 통한 스타트업과의 협력, 직접 투자(SI) 등을 집행해오고 있어 후속 투자 연계 등에 있어 협력 예정
	Pavillion Capital	싱가포르	• 현재 국내 VC 등에 펀드 출자 진행한 바 있으며 후기 단계의 경우 직접 투자함에 따라 스케일업을 위한 글로벌 투자 유치 시 연계 목표
	EDB	싱가포르	• 스타트업 현지 진출 시 법인 설립, 인재 채용 등에 대한 지원 제공 가능
	HTX	싱가포르	• 싱가포르 내부무 산하 기관으로(홈팀 과학기술청) 국가 및 사이버 보안 관련 스타트업 발굴 및 POC 협력 지원 가능
	Farquhar Venture Capital	싱가포르	• 국내 초기 스타트업에의 투자를 확대하고 있으며 공동 펀드 조성 등을 위한 MOU 체결 진행(23.12)하였으며 공동 투자 등에 있어서 협력 진행 예정
	Plug and Play	-	• PNP Korea 외에도 APAC, Japan, Thailand 등과 향후 국내외 유망 스타트업의 투자 딜을 공유 예정

24

© INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank

[연합뉴스] LG 그룹 혁신사례, MTV 5억불 매출 달성 시상 - 이한수 PBL

LG 스킬올림픽'
기사입력 2010-05-26 19:43
(https://m.blog.naver.com/hansleee/60108043714)
(서울=연합뉴스) LG가 25일과 26일 양일간 경기도 이천 LG인화원에서 경영혁신활동대회인 'LG 스킬올림픽'을 개최했다. 사진은 26일 구본무 LG회장이 우수혁신사례 영상을 보고 있는 모습. 2010.5.26 << LG >>



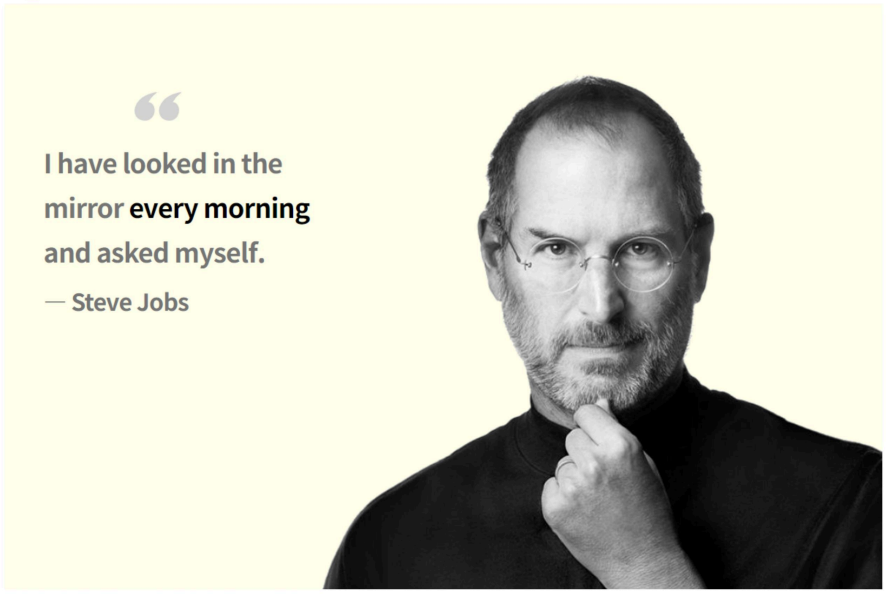


25

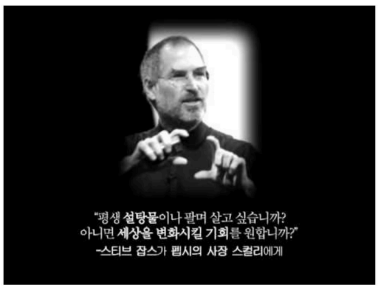
© INFOBANK. All right reserved.

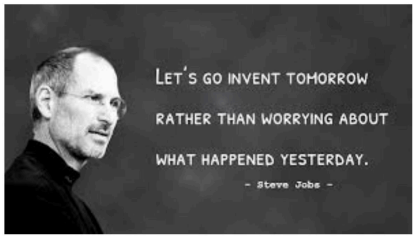
04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank



Stay Hungry와 Stay Happy





2011.10.5

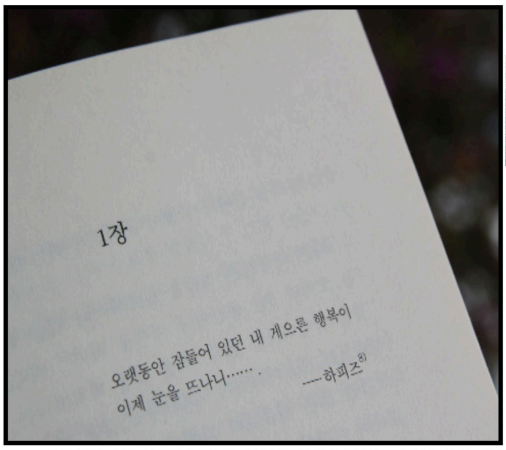
© INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank

사업과 행복





[THE FRUITS OF THE EARTH] ANDRE GIDE



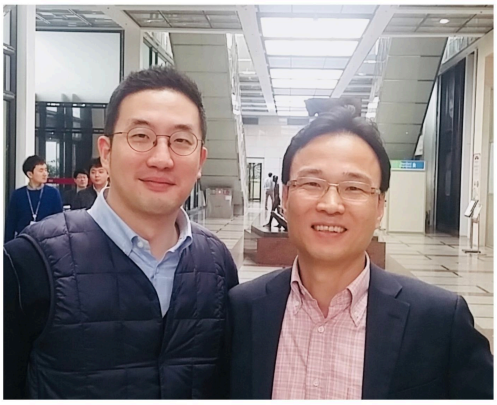
하와이 2015.8.2

© INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank

회장과 창업자



L그룹 회장님의 제안과 작별 2016.1.15



28

© INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank



S사 권오현부회장 + KSVF 스타트업 간담회, 2019.11.27

자립 / 자유 / 자족

세상을 바꾸는 것은 기업이고
기업은 창업(스타트업)의 결과이다.
그래서 스타트업이 미래다.

행복하게 살려면 자유가 있어야 하고
자유는 자립에서 나온다.
그런데 자족하지 않으면 불행해질 수 있다.

- 자립 자유 자족 -

29

© INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank



대왕과 철학자

디오게네스 친구 아리스티포스가 말했다.

"너가 왕(디오니시오스)에게 가서 머리를 조아리면
콩 같은 것을 먹지 않아도 될 텐데..."

그러자 디오게네스가 답했다.

"너가 콩을 먹는 것으로도 만족 한다면
왕에게 가서 머리를 조아리면서 살지 않아도 될 텐데..."

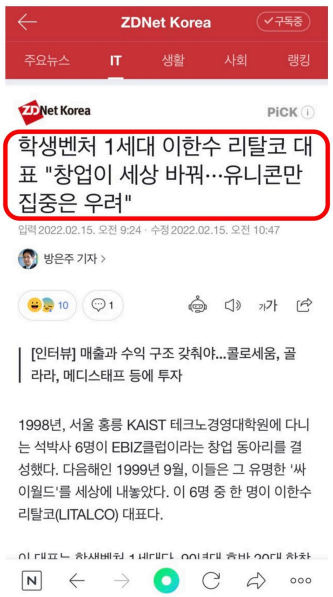
ALEXANDER THE GREAT VS. DIOGENES

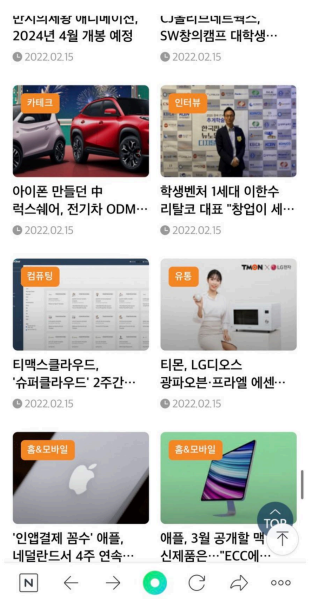
30

© INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank





31

© INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank





스포츠정신과 기업가정신

1. 배려와 매너
2. 열정과 노력
3. 승복과 존중



ⓒ INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank

전자신문 etnews

SW IT 경제 전자 모빌리티 플랫폼/유통 과학 정치 오피니언

SW > SW

인포뱅크, 이한수 리탈코 전 대표 영입…역량 강화

발행일 : 2024-03-18 09:15



이한수 인포뱅크 아이엑셀 부문 신임 파트너

인포뱅크는 투자 사업부 아이엑셀 역량 강화와 신규 성장동력 확보를 위해 이한수 리탈코 대표를 파트너로 영입했다고 18일 밝혔다.

Chosun

로그인

UPDATED. 2024-12-15 12:00 (일)

속보 기업 모빌리티 게임·콘텐츠 과학·헬스 파이낸스 칼럼·인터뷰 마소코딩 알림 GLOBAL

기업 > 일반

이한수 인포뱅크 파트너 "스타트업, 시드투자 의존 말고 계속 개발해야 성장"

김홍찬 기자

입력 2024.11.15 14:55수정 2024.11.15 17:35

"스펙만 화려하다고 좋은 스타트업이 아닙니다. 좋은 사업 모델을 바탕으로 계속 세상의 변화를 일으키는데 집중해야 합니다. 매출의 지속적인 성장과 함께 적정 시점에 이익도 내야합니다. 그래야 후속 투자와 성장으로 이어지는 겁니다. 우리는 그런 스타트업과 함께하기를 원합니다."



14일 서울 강남구 한국과학기술회관에서 열린 한국-스탠포드 벤처포럼(KSVF)에서 이한수 리탈코 대표 겸 인포뱅크 아이엑셀 파트너가 발표하고 있다. / 김홍찬 기자

이한수 인포뱅크 아이엑셀 파트너 겸 리탈코 대표는 14일 서울 강남구 한국과학기술회관에서 열린 한국-스탠포드 벤처포럼(KSVF)에서 '스타트업이 미래다!'를 주제로 발표하며 이 같이 말했다.

실시간 뉴스

- 11:15 스타벅스 "크리스마스에 해리포터 영화 함께 즐겨요"
- 11:11 이석용 농협은행장, 폭설 피해 농가서 "복구에 전사적 역량 결집"
- 10:47 현대자동차, '현대 N 페스티벌' 2024시즌 종합 시상식 개최
- 10:42 하나금융그룹, 취약계층 아동에 '행복상자' 전달
- 10:32 LG전자, AI 모듈러주택 '스마트코티지' 상용화…첫 고객은 SM엔터

시뮬레이션
해석 컨설팅
vpkcorp.com

ⓒ INFOBANK. All right reserved.

04 사업 성공만이 아닌, 스타트업과 함께하는 삶

infobank



[ZORBA THE GREEK] NIKOS KAZANTZAKIS

기업인과 자유인

나는 아무것도 바라지 않는다.

나는 아무것도 두려워하지 않는다.

나_는____자_유_다

[니코스 카잔차키스]

34

© INFOBANK. All right reserved.

스타트업과 함께 대한민국과 개인의 성공적인 미래를 준비하고,
자립을 통한 자유롭고 행복하고 만족스러운 삶을 만들어 가시길 기원드립니다~~~!!

Happy & Success : Happy HS

지상의 양식 - 앙드레지드
그리스인 조르바 - 니코스 카잔차키스
스티브잡스 - 월터 아이작슨
부자아빠 가난한아빠 - 로버트 기요사키
칼의 노래 - 김훈
여덟단어 - 박웅현
피로사회 – 한병철
코스모스 - 칼세이건
아름다운 마무리 - 법정스님

35


infobank
hanslee@infobank.net
infobank

Keynote Speech 02

Agentic AI 시대의 신뢰와 거버넌스 - 책임있는 AI로의 전환



공성배
메가존클라우드 부사장

발표 개요

Agentic AI는 단순한 자동화 수준을 넘어, 의사결정과 실행을 스스로 수행하는 지능형 에이전트(AI Agent)로 진화하고 있다. 이제 기업은 “AI를 어떻게 활용할 것인가”보다, “AI를 어떻게 신뢰하고 책임 있게 운영할 것인가”를 고민해야 하는 시점에 도달했다. AI가 자율적으로 판단하고 실행하는 시대에는 Digital Trust와 Governance Framework가 기술 혁신의 필수 전제조건이 된다.

이력

공성배 CAIO는 지난 수십 년간 여러 글로벌 기업에서 AI & 데이터 비즈니스를 담당해 온 AI & 데이터 분석 전문가이다. 기업이 데이터와 AI, 특히 생성형 AI를 비즈니스 전략의 중요한 요소로 삼고, 이를 통해 비즈니스 잠재력과 가치를 극대화할 수 있도록 조력하고 있다.

2018년 메가존클라우드에 합류해 AIR부문을 이끌며, 클라우드 기반 데이터 파이프라인 구축과 AI/ML 기반 분석 비즈니스를 전반적으로 총괄해 왔다. 이러한 경험을 바탕으로 2025년부터는 생성형 AI 전환을 위한 메가존클라우드의 전략적 AI 솔루션 통합브랜드 ‘MEGAZONE AIR’의 총괄 리더로서, 고객에게 실질적인 AI 활용 가치를 제공하고 실행 중심의 전환을 주도하고 있으며, 나아가 메가존클라우드의 AI 전략과 실행을 통합적으로 리드하고 있다. 조직 전반에 AI 기술을 확산시키고 혁신 문화를 내재화함으로써, 메가존클라우드가 단순한 클라우드 기업을 넘어 AI-Native Company로 도약할 수 있도록 변화의 방향을 이끌고 있다.

메가존클라우드에 합류하기 이전에는 한국테라데이타에서 컨설팅 부문 본부장(Professional Service Partner)으로 재임하며 Professional Service 비즈니스를 이끌었고, 14년간 IBM GBS(Global Business Services)에서 데이터 분석 컨설턴트로 활동하면서 국내 유수의 기업들이 AI와 데이터를 통해 비즈니스의 성공 지도를 그릴 수 있도록 기여했다.

	Memo	

Keynote Speech 03

무선백도어 해킹의 무한한 진화: AI시대 내부망 보호를 위한 제로 트러스트 전략



이원중
지슨 부사장

발표 개요

AI시대 보안 역시 AI에 의한 자동 모니터링 및 검증의 시대로 패러다임이 바뀌고 있다. 하지만 GPU·HBM 등 AI인프라의 코어 하드웨어 기반이 무선백도어 해킹에 의해 다층 공급망 단계에서부터 공격을 당한다면 AI의 디지털 신뢰에 대한 근본적인 문제가 발생하게 된다.

이번 키노트 세션에서는 무선백도어 해킹의 원리 및 사고 사례와 함께 주요국가들의 대응 방식에 대해 상세히 소개하고, 이에 대한 가장 현실적인 대안을 제시하고자 한다.

이력

- 고려대학교 경영학과 졸
- 국가정보원 근무(2004~2021)
- ㈜지슨 부사장(2022~현재)

무선백도어 해킹의 무한한 진화
: AI 시대 내부망 보호를 위한 제로 트러스트 전략

Copyright © 2025. GITSN, Inc. All rights reserved.

www.gitsn.com / sales@gitsn.com / 02.852.3560

목차

- 무선백도어 해킹의 개념 및 사례
- 무선백도어 해킹에 대한 대응
- 무선백도어 해킹 탐지 솔루션

무선백도어 해킹의 개념 및 사례

3

무선백도어 해킹의 무한한 진화 : **AI 시대 내부망 보호**를 위한 **제로 트러스트** 전략 GITSN(주)지슨

최근 발생한 주요 보안사고 4건 비교·분석

대상	A 통신사 (25.04)	B 보증 (25.07)	C 카드 (25.08)	D 통신사 (25.08)
공격방식	특정 신호(매직패킷)에만 반응하는 BPF 백도어 악용	외부 노출된 VPN 포트 대상 반복 로그인으로 초기 침투	웹셸(웹 서버에 업로드되어 원격 명령을 실행하는 악성 스크립트)	불법 펨토셀 장악 △BPF 백도어 △웹셸 악용

시그니처 기반 탐지 ⇒ 모니터링·행동 기반 탐지로의 패러다임 전환

4

APEC 기간 한중정상회담 관련 영상(채널A, 2025.11.01)



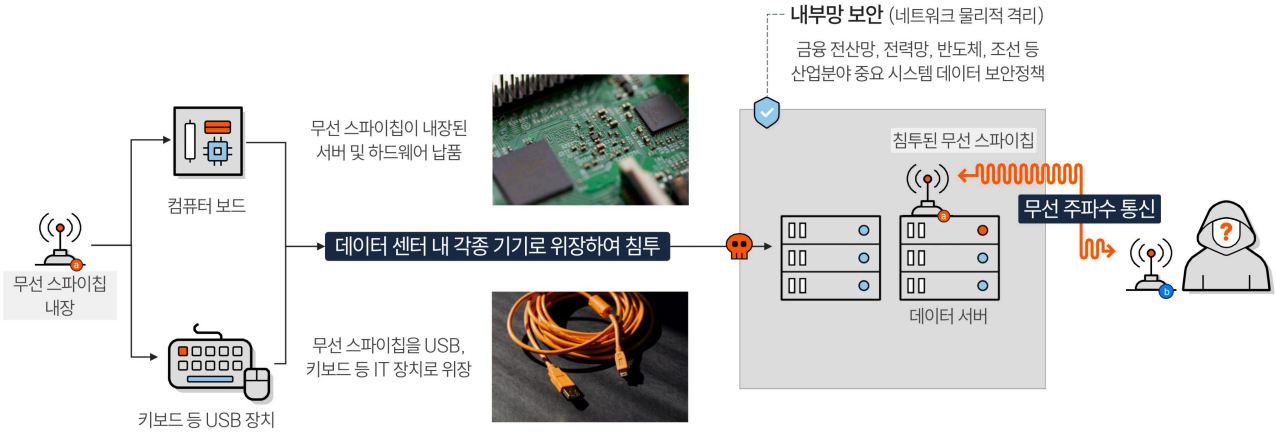
무선백도어의 정의 및 특징

① 무선 스파이칩(초소형 CPU 칩 + 무선통신 모듈)을 공급망 과정 중 메인보드에 탑재 또는 IT장치(키보드, USB 장치 등) 속에 은닉

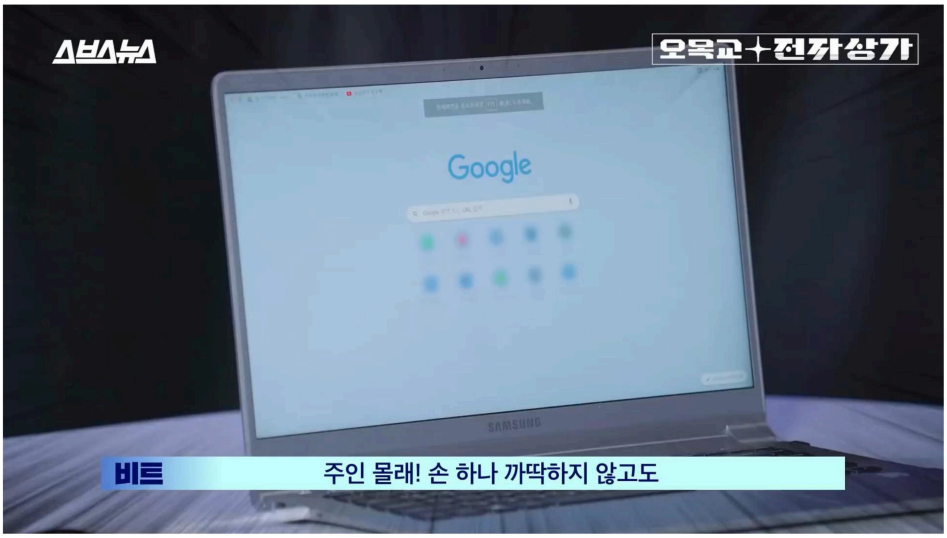
② 무선 주파수 통신으로 내부망으로 분리된 서버 등 타겟 시스템에 원격 접속(수 km밖에서 무단 침입)하여 데이터 탈취 및 시스템 붕괴

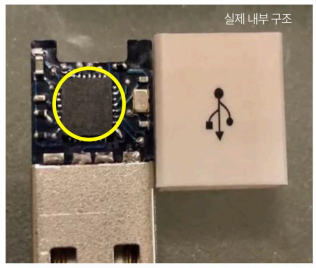
③ 기존의 보안 솔루션으로는 방어 불가능한 신종 해킹 **망분리 시스템(내부망) 무력화** **WIPS, 방화벽 등 기존 보안체계 우회**

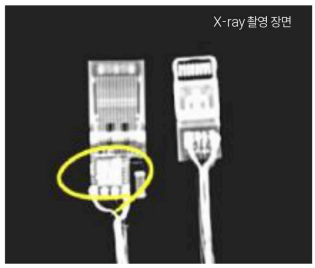
④ 매우 높은 수준의 접속 권한 획득 - 고부가가치 데이터 탈취, 사이버 테러로 인한 혼란 우려



'해킹 칩'을 숨겨 충전 케이블로 위장한 무선백도어 스파이칩 언론 보도(스브스뉴스, 2021.10.05)





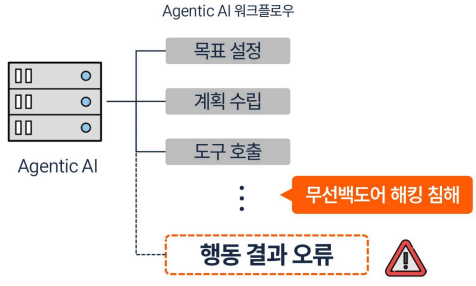


AI 시대 내부망을 노리는 가장 최신의 위협, 무선백도어 해킹

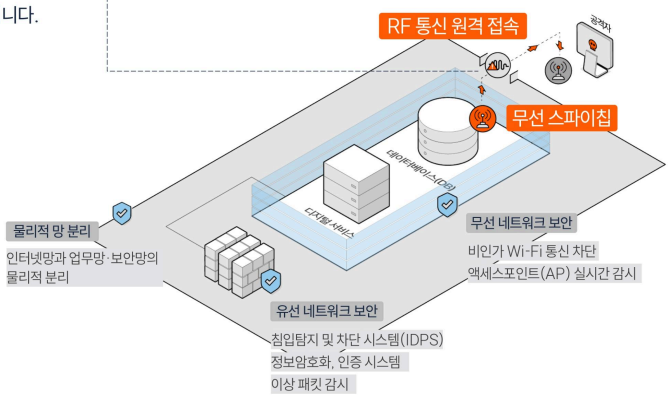
Agentic AI 시대에 더 위협적인 무선백도어 해킹

워크플로우 중 단 한 단계라도 하드웨어 백도어(스파이칩) 침해 발생 시

스스로 결정하고 실행하는 주체인 Agentic AI 전체의 신뢰 붕괴로 이어집니다.



무선백도어 해킹 통신 모듈과 초소형 CPU 칩으로 구성된 무선 스파이칩을 활용한 해킹으로, 공급망 해킹 또는 HW 백도어 해킹으로도 불림

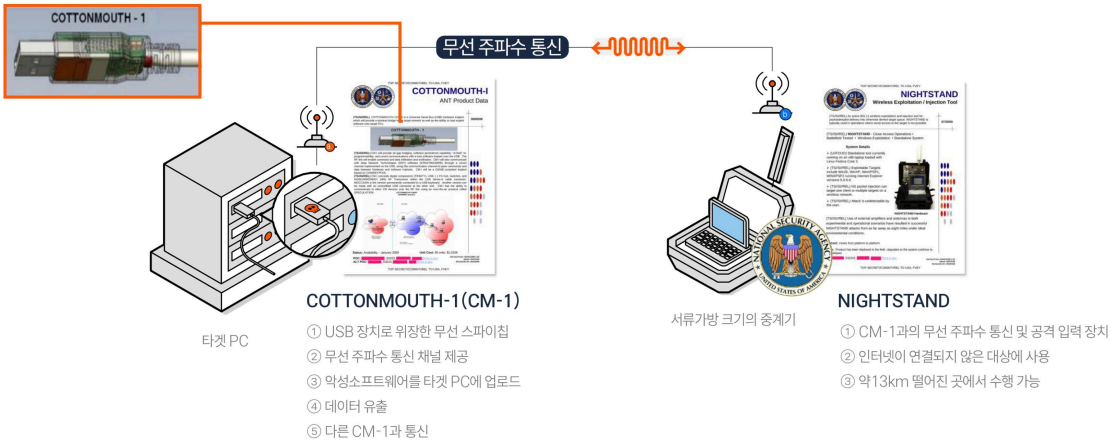


백도어 사례 (USB 장치 : 미국 국가안보국 퀀텀 프로그램)

① 미국 국가안보국(NSA)이 전 세계 10만대의 PC에 소프트웨어를 심어 정보를 빼내거나 사이버 공격에 활용(뉴욕타임즈 2014.01.14 보도)

② USB 장치(COTTONMOUTH-1)를 통해 타겟 PC안에 악성소프트웨어를 업로드, 장치가 발산하는 무선 주파수(RF)로 무선백도어 실행 (8 mile,13 km)

③ 중국 해킹부서, 러시아 군, EU 무역 담당 부처, 멕시코 경찰, 사우디아라비아, 인도, 파키스탄 등의 컴퓨터 네트워크에 설치



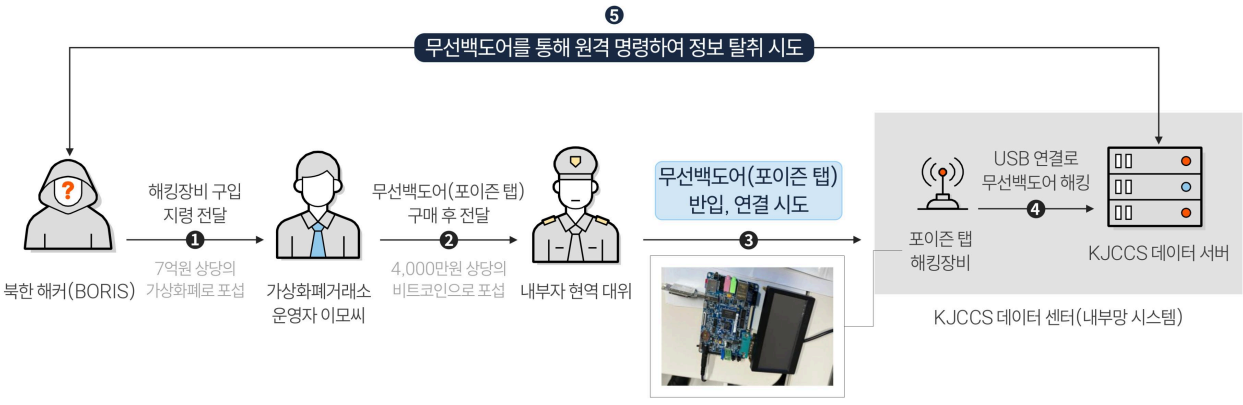
8

백도어 사례 (USB 장치 : 한국군 합동지휘통제체계 서버 해킹 시도)

① 현역 대위가 북한 해커에게 한국군 합동지휘통제체계(KJCCS) 로그인 자료 등을 제공 후 비트코인 수수(매일경제 2022.04.28 보도)

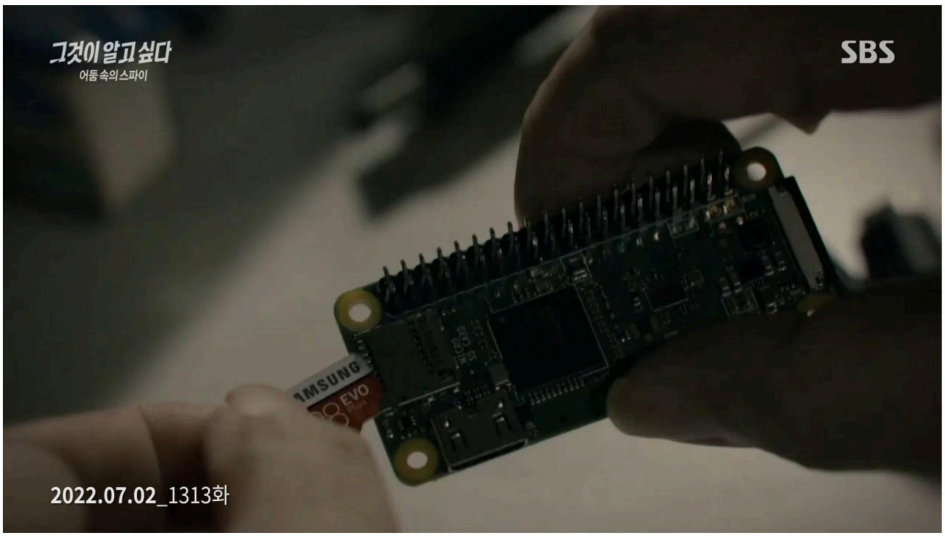
② 가상화폐거래소 운영자 이모씨가 군사기밀 탈취에 사용되는 무선백도어 장치(포이즌 탭) 구매 후 현역 대위에 전달

③ 무선백도어 장치(포이즌 탭)를 타겟 PC에 연결하면 북한 해커가 원격으로 무선백도어 해킹 가능



9

<그것이 알고 싶다> 1313화 포이즌탭 (SBS 그것이 알고 싶다, 2022.07.02)



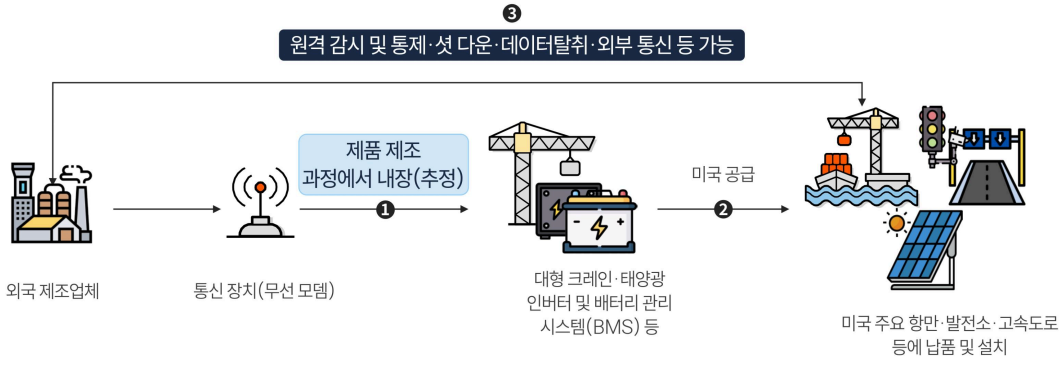
백도어 사례 (공급망, 보드: 외산 크레인·태양광 인버터 및 배터리 장비 등에서 비인가 통신 장치 발견)

① 美 항구 내 외산 크레인에서 요청하지 않은 통신장치 발견(2024.03.07 월스트리트 저널)

② 美 에너지부, 제조사 미상 외산 태양광 인버터에서 제품 설명서에 명시되지 않은 통신장치 발견(2025.05.14 로이터)

③ 美 연방고속도로청(FHWA), 제조사 미상 외산 태양광 배터리 관리 시스템(BMS)·인버터 내부에서 문서화되지 않은 통신장치 발견(2025.09.10 로이터)

④ 이러한 비인가 통신 장치에 대해 美 정부 측은 "데이터 탈취·원격 작동·설정 변경·셧 다운 등의 공격으로 인프라 교란 위험성이 있다"고 경고
⇒ 비인가 통신 장치 발견된 장비 즉각 퇴출 후, 정부 차원에서 정기적인 스펙트럼 탐지·분석, 발견 시 즉각 제거 등의 조치 권고



Agentic AI 시대의 무선백도어 해킹 공격 양상(예시)

무선백도어 해킹 공격은 다양한 보안기술이 적용된 공간에 침투하여 **수십 킬로미터 떨어진 거리에서도 해킹을 통한 정보 탈취 및 시스템 파괴** 가능

특히 AI가 자동으로 판단·실행하는 환경에서는 **서버 및 기타 하드웨어 무결성 침해가 곧 자동화 오작동으로 직결**되며,
인간이 중간에 개입하여 행동 오류를 차단·수정할 기회가 사실상 없기 때문에 이로 인한 경제적 손실·물리적 피해 발생 가능성이 높음

① 센서 데이터 위·변조(Sensor Spoofing)로 인한 물리적 손상

각종 센서에서 수집되는 온도·습도·압력·위치 등의 데이터를 위·변조하여 에이전트의 잘못된 제어 명령 혹은 AI 데이터센터 자체의 전력 과부하를 유도
⇒ 냉각 시스템 정지, 밸브 오작동 등 인프라 설비·장치에 대한 물리적 손상으로 직결

② 학습 데이터·인증정보 및 취약점 유출

개인정보·기밀 데이터·인증 정보·취약점 등이 외부로 유출 ⇒ 유출된 데이터로 인한 사회·경제적 손실
관리자 권한 탈취를 통해 클라우드 혹은 디바이스 내부망에 추가 침투하고 의도된 오출력을 유도해 AI 모델 학습을 오염시키는 등 무선백도어를 병행한 콤비네이션 해킹 가능

* 콤비네이션 해킹(Combination hacking) : 여러 종류의 해킹 기법이나 공격 벡터를 조합해서 사용하는 복합적인 해킹 방식으로, 단일 기법으로 뚫기 어려운 보안 체계를 다양한 방식의 공격을 동시에 혹은 순차적으로 사용해 우회하거나 무력화하는 전략

③ 서비스 정지·셧 다운(Shut down)

운영체계 변경, 보안설정 약화에 그치지 않고 특정 장비에 원격으로 셧 다운 명령 ⇒ 생산 라인·모니터링 작업 등의 연쇄 중단 발생
AI 모델이 가동되고 있는 물리 서버 중 일부 셧 다운 시에는 AI 모델의 원활한 작동 자체가 불가능하여 막대한 피해 발생

무선백도어 해킹에 대한 대응

세계 각국(미국, 일본, 한국)의 무선백도어 해킹 대응 방식

국가	대응 방식	특징
미국	전량 퇴출 및 자국산 장비로 교체, 향후 도입 배제 (국가안보법(2019) 개정 → 외산 장비 신규 납품 금지)	하드웨어 레벨 리스크 완전 제거 전략 장점: 공급망 레벨에서 잠재 위협을 원천 제거 단점 : △막대한 교체 비용 △기존 인프라 마이그레이션 시간 소요 △ 내·외부자의 사후 침투에는 무방비
일본	도시바, IDC 서버 보드 정밀 비파괴 검사 서비스 출시 (1장당 약 2만 달러, 1개월 소요)	설비 검증 기반 탐지 전략 장점 : 하드웨어 단계에 삽입되어 소프트웨어 보안 점검으로 식별하기 어려운 스파이칩을 물리적 검증을 통해 색출 단점 : △전수조사에는 부적합(비용·시간 부담) △원설계도 확보의 어려움 △ 내·외부자의 사후 침투에는 무방비
한국	**이상 주파수 실시간 탐지 시스템 도입,** 즉각 탐지·색출 후 제거	**신속·가성비 중심 대응 전략** 장점 : 무선백도어 해킹에 사용되는 이상 주파수를 최단 시간에 탐지하는 솔루션 도입으로, 시간·비용 측면에서 효율적 단점 : 무선 스파이칩 탐지 후 최종 제거 시에는 별도 인력 투입 필요

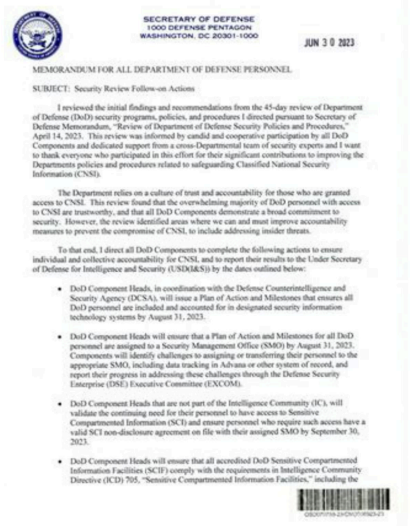
14

무선백도어 해킹의 무한한 진화 : AI 시대 내부망 보호를 위한 제로 트러스트 전략 GITSN(주)지슨

미 국방부 장관 메모(2023.06.30)

미 국방부는 SCIF/SAPF 등 기밀구역에 무선(전자기) 탐지시스템(WIDS/WRIDS)도입을 2024-09-30까지 후속 프로그램에 반영하도록 지시

SCIF : 민감 특수정보시설, Sensitive Compartmented Information Facility / SAPF: 특수 접근 프로그램 시설, Special Access Program Facility



SECRETARY OF DEFENSE
1000 DEFENSE PENTAGON
WASHINGTON, DC 20301-1000

JUN 30 2023

MEMORANDUM FOR ALL DEPARTMENT OF DEFENSE PERSONNEL

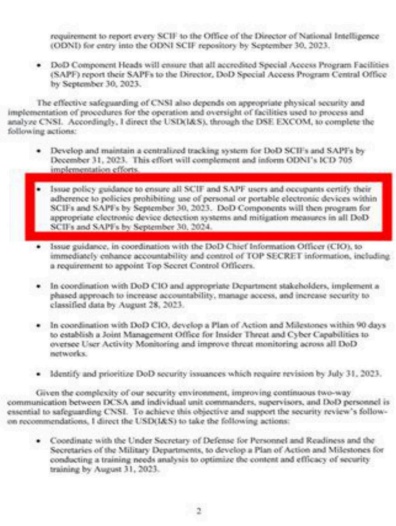
SUBJECT: Security Review Follow-on Actions

I reviewed the initial findings and recommendations from the 45-day review of Department of Defense (DoD) security programs, policies, and procedures I directed pursuant to Secretary of Defense Memorandum, "Review of Department of Defense Security Policies and Procedures," April 14, 2023. This review was informed by candid and cooperative participation by all DoD Components and dedicated support from a cross-Departmental team of security experts and I want to thank everyone who participated in this effort for their significant contributions to improving the Departments policies and procedures related to safeguarding Classified National Security Information (CNSI).

The Department relies on a culture of trust and accountability for those who are granted access to CNSI. This review found that the overwhelming majority of DoD personnel with access to CNSI are trustworthy, and that all DoD Components demonstrate a broad commitment to security. However, the review identified areas where we can and must improve accountability measures to prevent the compromise of CNSI, to include addressing insider threats.

To that end, I direct all DoD Components to complete the following actions to ensure individual and collective accountability for CNSI, and to report their results to the Under Secretary of Defense for Intelligence and Security (USD(I&S)) by the dates outlined below:

- DoD Component Heads, in coordination with the Defense Counterintelligence and Security Agency (DCSA), will issue a Plan of Action and Milestones that ensures all DoD personnel are included and accounted for in designated security information technology systems by August 31, 2023.
- DoD Component Heads will ensure that a Plan of Action and Milestones for all DoD personnel are assigned to a Security Management Office (SMO) by August 31, 2023. Components will identify challenges to assigning or transferring their personnel to the appropriate SMO, including data tracking in Advana or other system of record, and report their progress in addressing these challenges through the Defense Security Enterprise (DSE) Executive Committee (EXCOM).
- DoD Component Heads that are not part of the Intelligence Community (IC), will validate the continuing need for their personnel to have access to Sensitive Compartmented Information (SCI) and ensure personnel who require such access have a valid SCI non-disclosure agreement on file with their assigned SMO by September 30, 2023.
- DoD Component Heads will ensure that all accredited DoD Sensitive Compartmented Information Facilities (SCIF) comply with the requirements in Intelligence Community Directive (ICD) 705, "Sensitive Compartmented Information Facilities," including the



requirement to report every SCIF to the Office of the Director of National Intelligence (ODNI) for entry into the ODNI SCIF repository by September 30, 2023.

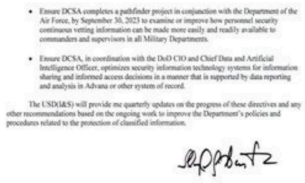
- DoD Component Heads will ensure that all accredited Special Access Program Facilities (SAPF) report their SAPFs to the Director, DoD Special Access Program Central Office by September 30, 2023.

The effective safeguarding of CNSI also depends on appropriate physical security and implementation of procedures for the operation and oversight of facilities used to process and analyze CNSI. Accordingly, I direct the USD(I&S), through the DSE EXCOM, to complete the following actions:

- Develop and maintain a centralized tracking system for DoD SCIFs and SAPFs by December 31, 2023. This effort will complement and inform ODNI's ICD 705 implementation efforts.
- Issue policy guidance to ensure all SCIF and SAPF users and occupants certify their adherence to policies prohibiting use of personal or portable electronic devices within SCIFs and SAPFs by September 30, 2023. DoD Components will then program for appropriate electronic device detection systems and mitigation measures in all DoD SCIFs and SAPFs by September 30, 2024.
- Issue guidance, in coordination with the DoD Chief Information Officer (CIO), to immediately enhance accountability and control of TOP SECRET information, including a requirement to appoint Top Secret Control Officers.
- In coordination with DoD CIO and appropriate Department stakeholders, implement a phased approach to increase accountability, manage access, and increase security to classified data by August 28, 2023.
- In coordination with DoD CIO, develop a Plan of Action and Milestones within 90 days to establish a Joint Management Office for Insider Threat and Cyber Capabilities to oversee User Activity Monitoring and improve threat monitoring across all DoD networks.
- Identify and prioritize DoD security issuances which require revision by July 31, 2023.

Given the complexity of our security environment, improving continuous two-way communication between DCSA and individual unit commanders, supervisors, and DoD personnel is essential to safeguarding CNSI. To achieve this objective and support the security review's follow-on recommendations, I direct the USD(I&S) to take the following actions:

- Coordinate with the Under Secretary of Defense for Personnel and Readiness and the Secretaries of the Military Departments, to develop a Plan of Action and Milestones for conducting a training needs analysis to optimize the content and efficacy of security training by August 31, 2023.



- Ensure DCSA completes a pathfinder project in conjunction with the Department of the Air Force, by September 30, 2023 to examine or improve how personnel security continuous vetting information can be made more easily and readily available to commanders and supervisors in all Military Departments.
- Ensure DCSA, in coordination with the DoD CIO and Chief Data and Artificial Intelligence Officer, optimizes security information technology systems for information sharing and informed access decisions in a manner that is supported by data reporting and analysis in Advana or other system of record.

The USD(I&S) will provide me quarterly updates on the progress of these directives and any other recommendations based on the ongoing work to improve the Department's policies and procedures related to the protection of classified information.

15

FHWA(美 연방도로청)가 각 주 정부 및 교통 당국에 배포한 권고사항(2025.08.20 배포)

분류	상세 내용
1. 자산 목록화 (Inventory)	- 전국적으로 설치된 태양광 배터리 및 관련 장비의 위치·제조사·모델·통신 기능의 전수 조사 지시 - '무엇이 어디에 있는지 모르면 방어할 수 없다'는 기본 대응 원칙 강조
2. 무선 신호 탐지 (Spectrum Analysis)	- **스펙트럼 분석기**를 이용해 **비인가된 셀룰러·무선 신호**가 송신되고 있는지 확인 조치 - 장비의 **정상 주파수 외에서 '의심스러운 송신 패턴'을 탐지**하라는 구체적·기술적 지시
3. 의심 장비 제거 (Remove any such devices)	- **문서화되지 않은 RF 통신 모듈**이 실제로 발견되면 **즉시 제거하거나 장비 교체** 지시 - 비활성화로는 불충분하며 완전한 색출 및 제거 권장
4. 네트워크 분리·보안 강화 (Network Security / Segmentation)	- 장비가 연결된 제어망·통신망을 구분하여 비인가 통신이 전체 인프라로 확산되지 않도록 구역화(segmentation) 강화
5. 제조·조달 검증 (Procurement Controls)	- 특정 제조국을 명시하지는 않았으나, 외국산 장비 반복 언급을 통해 **조달 단계에서의 검증 강화** 필요성을 시사

무선백도어 공격에 대한 방어대책 촉구(국정감사)

무선백도어 공격에 대한 대비로 금융권 정보 탈취 사전 방지 시스템 주문

국정감사 '정무위원회' 금융감독원 감사中 (2019.10.08)

김선동 의원 국내 은행하고 증권사의 무선 네트워크 정보 탈취 이런 부분들을 **무선백도어 침입**이라고 편의적으로 말씀드릴 수 있는데 …
내부 전산실 침입을 하려고 그러면 **스파이칩을 심거나 스파이칩이 내장된 전산장비를 설치하면 무방비 상태**로 뚫립니다. … 스파이칩이 내장된 키보드나 마우스 같은 것만 교체를 해도 이것이 뚫리게 되어 있습니다…

윤석현 금융감독원장 예, 저희들이 잘 모니터링을 하도록 하겠습니다.

군 내부망 해킹 보완을 위해 무선 통신 가능 전 주파수 대역 보안 시스템 촉구

국정감사 '국방위원회' 국방부 감사中 (2019.10.21)

백승주 의원 국방부의 군 내부망 해킹에 조금 새로 보완할 요소가 있다고 듣고 있습니다.
무슨 내용인지 아시지요?
무선통신 가능 주파수 전 대역에서 이 보안 시스템을 확인해야 될 부분이 있다.
그것 좀 확인해 주시고.

정경두 국방부장관 예. 알겠습니다.

금융권 무선백도어 해킹 공격 관련 질의 및 대비책 마련 촉구

국정감사 '정무위원회' 금융감독원 감사中 (2024.10.17)

강명구 의원 … 최근에는 **망분리 상태에서도 해킹이 가능한 무선백도어 해킹**이라는 게 늘어난다 그래요.
이게 예전에는 소프트웨어 통해가지고 백도어 해킹을 했다면 **이제는 아예 그냥 전자기기에 무선 스파이칩을 심어가지고 해킹하는 백도어 해킹에 대해 경각심**이 높아지고 있다는데,
금융기관들이 무선백도어 해킹도 금융보안 범죄에 포함시키고 있는 거죠?

이복현 금융감독원장 예, 지금 계속 그 신종 수법이라든가 … 저희 금융보안원 그리고 각 금융회사들이 계속 방어수단을 마련하려고 노력 중입니다.

무선백도어 해킹 관련 질의 및 ISMS 인증 항목 검토 등 대비책 마련 촉구

국정감사 '과학기술정보방송통신위원회' 과학기술정보통신부 감사中 (2025.10.13)

조인철 의원 … 공동으로 우리가 쓰고 있는 이 컴퓨터를 한 번 뚫고 들어가 보겠습니다. **(무선백도어 해킹 시연) 지금 컴퓨터 공동망(내부망)에** … 너무나 간단합니다. …**온라인으로 미국에 주문**을 했는데 **누구도 제지하지 않고 통관절차 정식으로 거쳐서** 들어왔습니다. **단돈 20만 원**도 안 합니다. …과기부에서 하고 있는 **ISMS 인증 제도** 있죠. 거기 항목들을 봤더니 여기**(무선백도어 해킹)에 대한 대책 항목은 하나도 없습니다.** 항목, 혹시 있다고 생각하십니까? 이거 체크하나요?

배경훈 과기부장관 저희가 80개 항목 중 스파이칩 관련된 거는 확인을 해 봐야겠습니다만, 명확하게 그렇게 구체적인 것(항목)은 없었던 것 같습니다.

무선백도어 공격에 대한 방어대책 촉구(규제 및 법률)

국가 새 사이버보안체계, HW 보안성도 검증 (2024.10.24, 전자신문)

25년 시행 목표인 국정원 다층보안체계(MLS)의 보안통제 항목에 '하드웨어 자재명세서(H-BOM)' 포함 방안 유력 검토 중

… 통제 항목에 H-BOM을 담으려는 것은 **하드웨어가 보안 통제 사각지대에 놓였다는 판단**에서다. 현재 서버와 스토리지 등 하드웨어는 보안 검증 없이 운영성 관련 인증만 거치면 국가·공공기관에 납품되고 있다.

하드웨어에 스파이칩 등 악의적 부품 탑재 여부, 외부에서 접속 가능한 백도어(인증 절차 없이 시스템에 접근할 수 있도록 만들어 놓은 비공개 통로) 검증 절차 등이 없다. 서버에 탑재되는 애플리케이션이 보안적합성 검증 대상 제품일 경우에도 애플리케이션만 검증받고 서버는 무사통과다.

… "**악의적인 스파이칩을 설치해 데이터 센터에 화재를 발생시키거나 전체 시스템 전원을 동시에 온·오프해 과부하를 일으키는 공격은 일도 아니다**"라며 "정부가 하드웨어를 철저히 검증하거나 개발사가 하드웨어 무결성을 입증하는 제도적 장치가 필요하다"고 강조했다. …

「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 일부 개정 (2024.1.23)

제48조(정보통신망 침해행위 등의 금지)

① 누구든지 정당한 접근권한 없이 또는 허용된 접근권한을 넘어 정보통신망에 침입하여서는 아니 된다.

② ~ ③ (생 략)

④ 누구든지 정당한 사유 없이 정보통신망의 정상적인 **보호 · 인증 절차를 우회**하여 정보통신망에 **접근할 수 있도록 하는 프로그램이나 기술적 장치** 등을 정보통신망 또는 이와 관련된 정보시스템에 설치하거나 이를 전달 · 유포하여서는 아니 된다.

18

무선백도어 해킹 보안 솔루션

19

무선백도어 해킹 기술의 발전, 더 이상 안전지대가 아닌 내부망

무선백도어 해킹 공격은 기존 경계 기반 보안 시스템에서는 탐지조차 어려워 **제로 트러스트로의 보안 패러다임 전환**이 필수
외부와 통신하지 않는 On Device AI의 경우라도 하드웨어 백도어를 통해 침투하는 무선백도어 해킹에는 무방비
⇒ 무선백도어 해킹이 가능한 전 주파수 대역에 대해 지속적인 감시와 검증 필요

① Wi-Fi 이외 대역의 주파수 사용

무선백도어 해킹을 시도하는 해커들은 **Wi-Fi 외 비표준 무선통신의 주파수, 변조 방식, 프로토콜 구조**를 원하는 대로 설계하여 이용
Wi-Fi를 통한 침입만을 탐지하고 프로토콜 기반 통신 차단 방식을 이용하는 **WIPS로는 탐지 불가**

② 네트워크 로그 기록이 남지 않음

스파이칩을 통한 RF 통신은 내부 네트워크를 거치지 않고 외부 페어 칩과 1:1 로 이뤄지기 때문에 **침입 흔적이 남지 않음**
기존의 네트워크 로그 감시체계를 우회하는 통신 방식으로, 전대역 주파수 스캔이 아니면 비인가 접근이나 데이터 유출을 탐지할 수 없음

③ 납품 장비의 사전 검수 어려움

현실적으로 납품되는 모든 서버와 IT기기를 **원설계도면과 비교하여 검수하기란 불가능**에 가까우며, **납품 시점에 스파이칩이 비활성 상태일 경우 사전 탐지도 불가**

④ APT 공격 증가로 장시간 잠복

APT 공격에서 **무선백도어 해킹은 이미 강력한 공격 수단으로** 사용되며, 온라인상에서도 관련 해킹 기술이 많이 공개됨
긴 시간 신호 송수신 없이 잠복해 있다가 최적의 타이밍에 공격을 개시하는 방법을 사용하는 경우 사전 탐지가 어려움

무선백도어 해킹 탐지 시스템의 필수 요건



무선 스파이칩 전파 탐지

25kHz~6GHz의 주파수 대역을 스캔하여 무선백도어 해킹 탐지



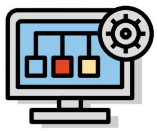
24시간 상시 탐지

보안의 빈틈을 없애는 24시간 365일 상시 탐지



위치 추정

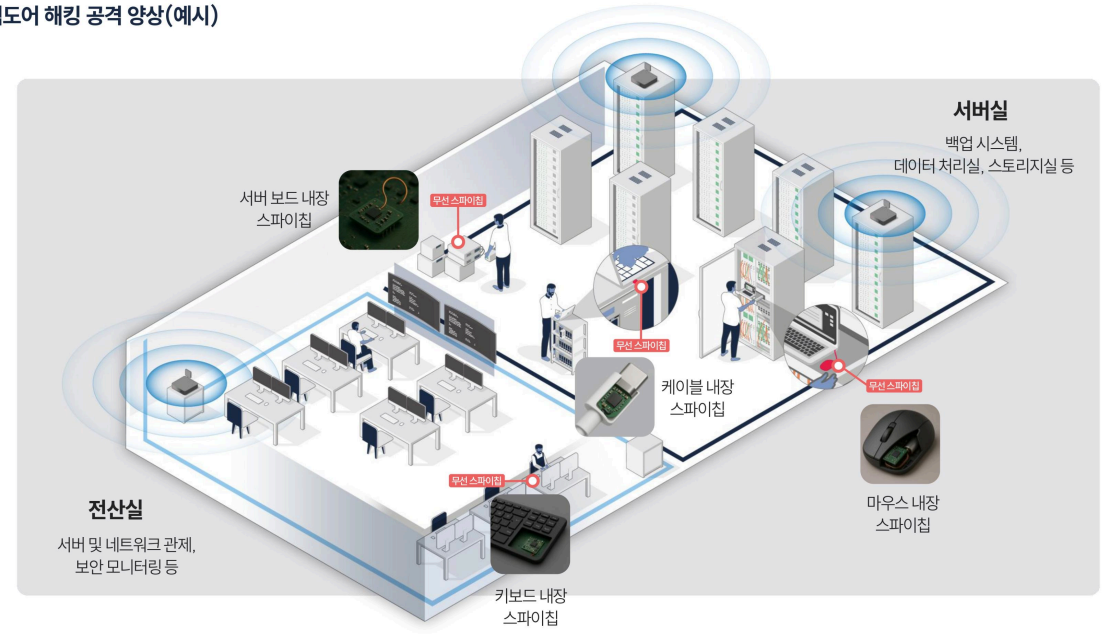
이상 주파수의 신호원 위치 추정으로 무선백도어 해킹에 신속 대응



통합 관제

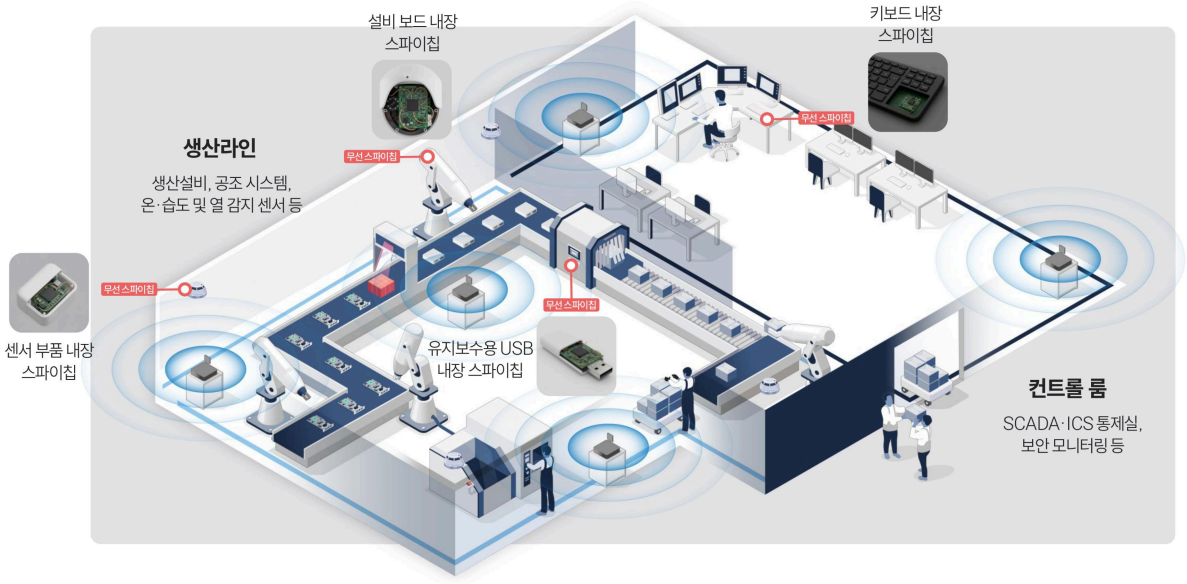
다수 탐지단말기 통합 관리·관제

무선백도어 해킹 공격 양상(예시)



무선백도어 해킹의 무한한 진화 : AI 시대 내부망 보호를 위한 제로 트러스트 전략 GITSN (주)지슨

무선백도어 해킹 공격 양상(예시)



결론

24

결론



"반응의 시대를 지나 예측의 시대로"

보안은 '사고 이후' 대응이 아니라 '가능성을 읽는' 상상력에서 시작됩니다.

상상력의 부재가 불러오는 재앙을 미리 대비해야 합니다.

25

Keynote Speech 04

AI에게는 난해한 인간의 감정(행복)



서은국
연세대학교 교수

발표 개요

AI와 인간의 교류는 점점 가속화될 것이다. 하지만 모방이나 통계적 추론은 가능해도 AI 입장에서 인간에 대해 제대로 “이해”하기 어려운 특성들이 있을 것으로 보인다. 대표적으로, 왜 인간은 행복과 같은 감정적 경험을 그토록 중시할까? 강의를 통해 최근의 진화심리학 관점으로 재해석하는 행복(감정)의 기능과 존재 이유를 소개한다. AI와 더불어 사는 시대에 여러 기술적, 철학적 함의점이 담긴 질문이다.

이력

연세대학교 졸업 후 미국 일리노이 대학교에서 행복 분야 권위자인 에드 디너 교수의 지도를 받고, 심리학 박사 학위를 취득했다. 학위 후 캘리포니아 주립 대학교에서 교수 생활을 시작했고, 4년 뒤 이 대학에서 종신 교수직을 받았다. 세계에서 가장 활발하게 인용되는 행복 심리학자 중 한 명으로, 특히 행복과 문화 차가 전문 분야다. 그의 논문들은 OECD 행복 측정 보고서에 참고자료로 사용되고 있으며, ‘세계 100인의 행복 학자’에 선정되어 『세상의 모든 행복에 기고했다.

	Memo	

2025 공동 학술대회

Track 1



Track 1-1

Agent AI에 관한 기본적 법률 이슈 검토요건

최주선
네플라 주식회사 CEO



Track 1-2

초소형 탈중앙화 고밀도 데이터센터와 온프러미스AI를 통한 혁신적인 인공지능 생태계 구축

조정현
주식회사 제타규프 대표



Track 1-3

불확실성 시대에서의 AI 거버넌스와 리스크관리 전략, 새로운 AI 감사 자격 체계 AAIA를 중심으로

류종기
EY한영 상무



Track 1-4

AI 시대의 안전한 의료 AI를 위한 책임과 이슈: 고영향 AI 확인 및 사업자 책무 가이드라인을 중심으로

유석
유니닥스 본부장



Track 1-5

Agentic AI 시대의 서비스 경험 혁신과 신뢰 거버넌스

김정훈
삼성전자 Principal Professional

	Memo	

Track 1-1

AI 에이전트에 관한 기본적 법률 이슈



최주선
네플라 주식회사 CEO

발표개요

내 컨디션을 분석해 월경시기를 예측하고 미리 생리대를 주문해 주는 AI 에이전트가 쿠팡에서 평소의 10배 수량의 생리대를 구매하면, 그 구매 계약의 효력은 어떻게 될까? 메타의 AI 에이전트가 내 인스타 계정을 정지해 버리면?

내 생활과 업무 전반에 파고들 AI 에이전트가 우리에게 미칠 법적 영향을 살펴보고, 이에 대해 우리는 이 문제들을 어떤 식으로 풀어갈 수 있을지 생각해 봅니다.

이력

- 서울대 철학과(법학 부전공)
- 카이스트 최고경영자과정(AIM) 수료
- 카이스트 CAIO과정 수료
- 사법시험 52회, 연수원 42기
- (현)네플라 주식회사 CEO
- (현)법무법인 민후 대표변호사
- (현)한국포스트휴먼연구소 이사장
- (현)서울지방변호사회 인공지능 및 리걸테크 대응 특별위원회 위원
- (현)법률신문, 패스트캠퍼스 법률분야 AI 활용법 강사
- (현)제6기 공공데이터전략위원회 위원
- (현)서울시 개인정보 보호 심의위원회 위원
- 개인정보 분야 대통령 표창

	Memo	

Track 1-2

초소형 탈중앙화 고밀도 데이터센터와 온프러미스AI를 통한 혁신적인 인공지능 생태계 구축



조정현
주식회사 제타큐브 대표이사

발표개요

인공지능시스템을 위한 인프라와 플렛폼의 혁신을 통해 AI를 빠르게 대중화 할 수 있는 대안을 제시 하려 합니다.

이력

- 1970년 서울 출생
- 한국항공대학교 정보공학 석사 졸업
- 1989년 프리렌서 소프트웨어-하드웨어 개발자
- 1996년 N.T!System 설립, iT 솔루션 개발
- 1997년 PC통신 오픈소스 동호회 회장 3년 역임
- 1999년 NeonGate Inc 로 법인 전환, 통신장비 제조, 수출
- 2006년 코발트레이 설립, 안드로이드 기반 모바일, 결제 장비 제조
- 2009년 안드로이드 협회 가입, 세계최초 양면 태블릿 출시
- 2022년 제타큐브 설립, 블록체인, 탈중앙화 데이터센터 솔루션 개발
- 2023년 초소형 탈중앙화 데이터센터 ‘NANODC’ 발표
- 2025년 ISMS, VASP 라이센스
- Filecoin 공인 Allocator, AI License 출제위원

초소형 탈중앙화 고밀도 데이터센터와 온프레미스 AI를 통한 혁신적인 인공지능 생태계 구축

주식회사
제타큐브 | 대표이사 조정현



Confidential. For internal use only.

ISACA Korea Chapter

회사 소개 초소형 탈중앙화 고밀도 A.I 데이터센터 솔루션 전문 기업

- 2022년 설립, 전국 5곳에 **탈중앙화 데이터센터** 구축
- 자체 **스토리지솔루션으로** Web3 탈중앙화 프로젝트에 참여
- 2023년 8월, 세계최초로 **NANODC(초소형데이터센터)**발표
- 2024년부터 **DePIN(탈중앙 물리 인프라 네트워크)** 개념 도입
- 세계 최대 **GPU 클라우드 DePIN 네트워크 Aethir**와 전략제휴
- 2025년 **ISMS 가상자산사업자 예비 인증 획득**
- **DePIN 생태계용 토큰 유동화 풀** 구축
- **AI 서버 32B, 70B, 671B 출시**
- **World iT Show 2024,2025** 참여
- DePINA Cohort 0 컨퍼런스 개최
- DePIN 협회 조직
- **2025 K-Digital Brand, 국가공헌 대상 수상**



2

대표 소개 조정현 (Allen Cho)

- 1970년 서울 출생
- 한국항공대학교 정보공학 석사 졸업
- 1983년 종로 세운상가 PC샵 근무
- 1989년 프리렌서 프로그램 개발 용역
- 1996년 N.T!System 설립, 솔루션 개발
- 1997년 PC통신 오픈소스 동호회 회장 3년 역임
- 1999년 NeonGate Inc 로 법인 전환
- 2002년 KDSL 22개국 수출, 580억원 매출
- 2006년 코발트레이 설립
- 2009년 안드로이드 협회 가입, 세계최초 양면 태블릿 출시
- 2022년 제타큐브 설립
- 2023년 NANODC 발표
- Filecoin 공인 Allocator, AI 인증시험 출제위원



3

핵심 사업 전략 인공지능 시대의 인프라 - 플랫폼 공급 혁신 기업

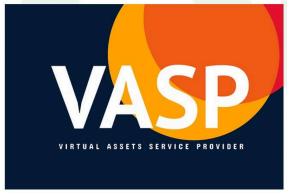
- ISMS/VASP 라이센스 기반의 초소형 탈중앙화 고밀도 데이터센터에 자체제작 AI서버로 AI 대중화 선도



NANODC







4

개발팀 소개 하드웨어-소프트웨어-블록체인-인공지능 개발 및 제조 전문가 그룹


제 2023112529 호

기업부설연구소 인정서

1. 연 구 소 명: (주)제타큐브 탈중앙화 스토리지 연구소
[소속기업명: (주)제타큐브]
2. 소 재 지: 서울특별시 서초구 논현로 63
6층 (양재동)
3. 신고 연월일: 2023년 05월 19일
※ 유효기간: 2025년 4월 19일(기업 인정일로부터 2년간)

과학기술정보통신부

「기초연구진흥 및 기술개발지원에 관한 법률」 제14조의 2제1항 및 같은 법 시행령 제27조제1항에 따라 위와 같이 기업부설연구소로 인정합니다.

2023년 5월 26일

한국산업기술진흥협회장

기술이사
정호석
- 캐나다 브리티시 컬럼비아 학사
- 스타트업 창업 개발 10년 경력
- 웹3, 게임 소프트웨어 전문가
- 국제 헤커톤 대회 8회 우승

대표 / 기술총괄
조정현
- 서울과기대 제어계측, 정보공학석사
- 30여년 IT 기업 개발 총괄 / 대표
- 소프트웨어 / 하드웨어 전문가
- DePIN 솔루션 연구개발 리드

사업본부장
이은정
- 마케팅 기획 및 디자인 전문가
- 10여년간 브랜딩 기획 종사
- iT제조,개발,인공지능 자격 보유
- 프로젝트 기획 관리 마케팅

연구팀장
최철훈
- 미 하와이대 컴퓨터사이언스 전공
- 소프트웨어 전문가
- 분산처리 소프트웨어 구성 담당

선임연구원
김대익
- 전북대 전자공학과 졸업
- 블록체인 개발 전문가
- DeFi, DAO, Dapp 개발담당

책임개발자
신예진
- 미 존 허시 스쿨 졸업
- 분산처리 스토리지 기업 근무
- 서버 시스템 구축 지원

선임연구원
조정빈
- 상지대학교 컴퓨터공학전공
- 소프트웨어 전문가
- 블록체인 스마트컨트랙트 담당

책임개발자
정의성
- 중부대학교 정보공학 졸업
- 네비게이션 시스템 개발
- APP, 프론트 개발

책임개발자
윤선화
- 대림대학교 광고홍보 졸업
- UX / UI 디자인
- WEB, APP, 프론트 퍼블리셔

주요 제품 소개 데이터센터부터 인공지능 서버, 운영 시스템 토탈 솔루션



NANODC

무정전장치가 결합된
특수 서버렉과
최적화된성능의 GPU,
스토리지서버로 구성된
초소형 탈중앙화 전용
데이터센터 솔루션



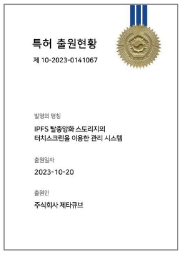
SAI
System to A.I Integration
systemto.ai

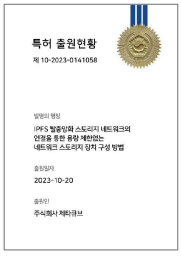
오픈소스 LLM 32B,
70B, 671B 를 독립
설치한 AI 엔진 서버로
VRAM 768G, 10560
TFLOPS의 강력한
컴퓨팅 파워를 자랑하는
하드웨어 아키텍쳐기반의
완전한 On-Premise AI
Server Solution



FILNAS

100테비바이트
저장이가능한
NAS서버로 초고속
40기가 통신을
바탕으로 타 기기로
백업이 매우 빠르며
컴팩트 하여 이동이
자유로운 네트웍
스토리지 백업장치


특허 출원현황
제 10-2023-0141067
IPFS 탈중앙화 스토리지의 터치스크린을 이용한 관리 시스템
출원일자
2023-10-20
출원인
주식회사 제타큐브


특허 출원현황
제 10-2023-0141058
IPFS 탈중앙화 스토리지 네트워크의 안정을 위한 유무 체인리는 네트워크 스토리지 장치 구성 방법
출원일자
2023-10-20
출원인
주식회사 제타큐브


특허 출원현황
제 10-2023-0120689
초소형 고성능 데이터센터 구축을 위한 서버랙 구성 방법
출원일자
2023-09-11
출원인
주식회사 제타큐브


실용신안 출원현황
제 20-2023-0001893
초소형 고성능 데이터센터용 서버랙
출원일자
2023-09-11
출원인
주식회사 제타큐브


상표등록 출원현황
상표명
NANODC
상품류
제09류
출원번호
40-2023-0174614
출원일자
2023-09-22
출원인
주식회사 제타큐브


상표등록 출원현황
상표명
ZETACUBE
상품류
제09류
출원번호
40-2023-0174622
상품류
제42류
출원번호
40-2023-0174534
상품류
제37류
출원번호
40-2023-0174544
상품류
제38류
출원번호
40-2023-0174603
출원일자
2023-09-22
출원인
주식회사 제타큐브

왜 AI에 탈중앙화 고밀도 데이터센터가 필요한가?

- 고전력이 필요한 고밀도 데이터센터는 구축이 어려워 수요를 충족 못하고 있습니다.

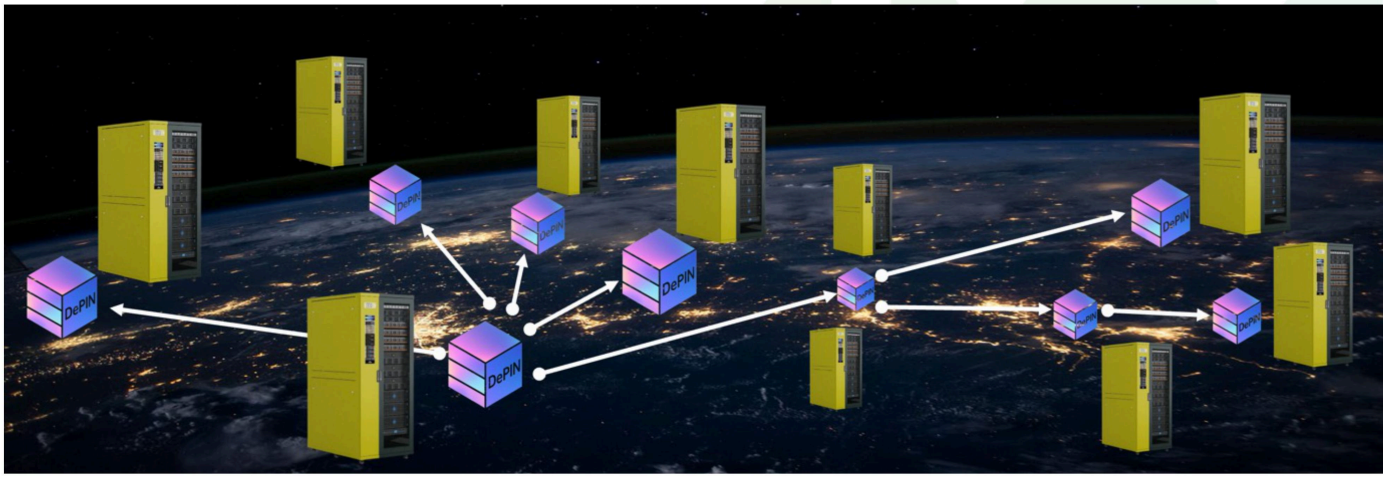


- 중앙집중화된 데이터센터로는 서비스, 데이터 안전을 보장 못합니다.

7

탈중앙화데이터센터는 지리적으로 분산되어야 합니다

- 네트워크 사용 속도, 고 전력 공급, 보안, 복구, 경제성 모두 유리합니다.



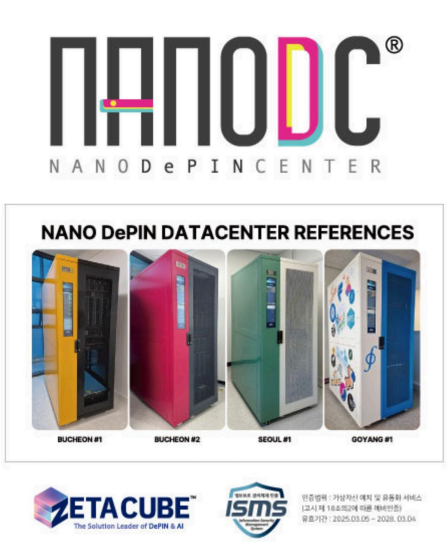
8

AI 인프라 구축 Killer Solution

- Nano Data Center 한대면 최대 20대의 AI 서버 수용 가능 (2U기준)



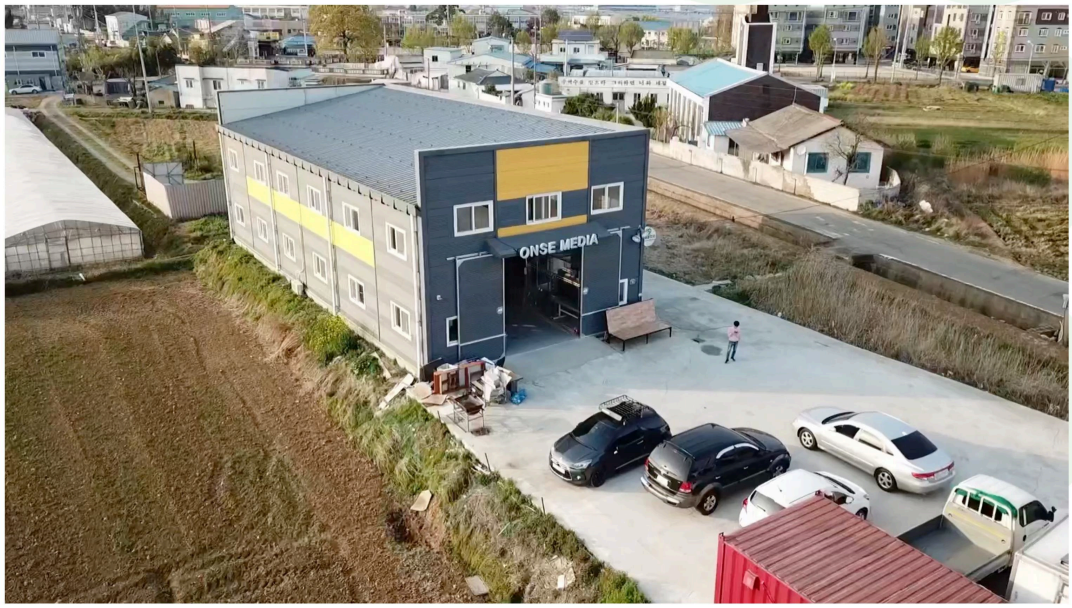




9

설치는 단 하루만에 가능

- 데이터센터 구축이 어려운 어떤 지역도 가능! (아프리카, 중동, 남미등)



10

단 세평의 공간만 있으면 어디든 데이터센터

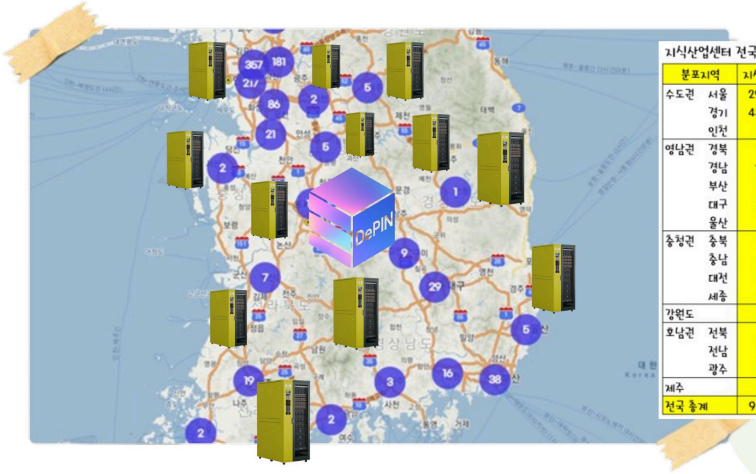
- 15KW산업용 전력 (최대40KW가능) 과 1G (최대 10G)인터넷 연결이면 끝!



11

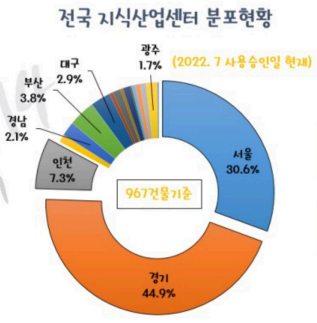
NANODC 보급 전략 지식산업센터와 소형 공장

- 저렴한 임대료 (월 50만원수준) 와 전력비용 (월 100만원 미만)이 장점



지식산업센터 전국 분포 및 일반건축물 비율

분포지역		지산수	점유율	일반	지역대비	집합	점유율
수도권	서울	296	30.6%	4	1.4%	292	32.1%
	경기	434	44.9%	8	1.8%	426	46.9%
	인천	71	7.3%	1	1.4%	70	7.7%
영남권	경북	11	1.1%	4	36.4%	7	0.8%
	경남	20	2.1%	6	30.0%	14	1.5%
	부산	37	3.8%	6	16.2%	31	3.4%
	대구	28	2.9%	7	25.0%	21	2.3%
	울산	5	0.5%	1	20.0%	4	0.4%
충청권	충북	11	1.1%	1	9.1%	10	1.1%
	충남	4	0.4%	2	50.0%	2	0.2%
	대전	10	1.0%	2	20.0%	8	0.9%
	세종	2	0.2%	1	50.0%	1	0.1%
강원도		5	0.5%	3	60.0%	2	0.2%
호남권	전북	7	0.7%	3	42.9%	4	0.4%
	전남	8	0.8%	4	50.0%	4	0.4%
	광주	16	1.7%	4	25.0%	12	1.3%
제주		2	0.2%	1	50.0%	1	0.1%
전국 총계		967	100.0%	58	6.0%	909	100.0%



전국 1,500 개 이상의 지식산업센터는 현제 공실률이 60%이상으로 임대료가 저렴하면서도 산업용 전력 사용을 통해 최대 40KW까지 고 전력 지원이 가능하며 산업용 전력으로 비용이 매우 저렴해 데이터센터 운영 수익을 극대화 할 수 있습니다.[12]

SAI (System to AI Integrate) 인공지능 시스템 통합 서비스

System to A.I Integration

systemto.ai

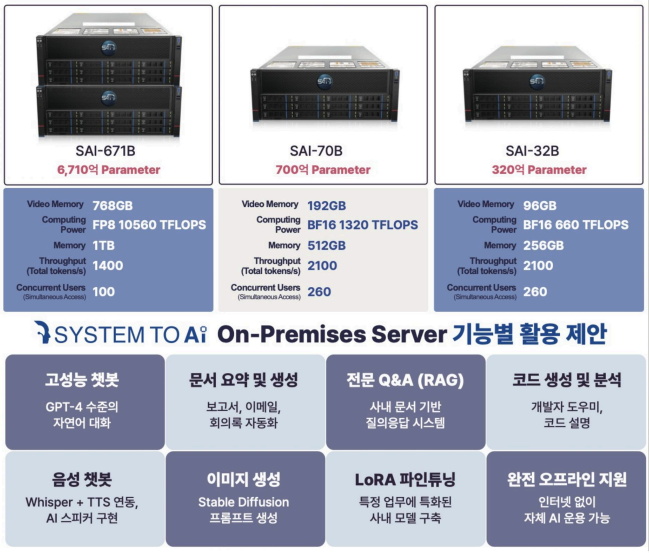
- 기존 시스템을 인공지능을 추가하여 통합하는 작업
- 내부 정보 유출 없이 내부자료로 인공지능 학습 가능
- 완전한 독립 설치가 가능해 완벽히 외부 침입 차단 가능
- 초거대 LLM 구동을 위한 시스템 (최대 671B 파라미터)
- 최적화 된 시스템 설계로 가격 대비 성능 매우 우수
- 문서, 이미지, 음성 등 다양한 데이터 처리 가능
- 고객 업무 인공지능화를 위한 개발 및 기술 지원
- 지속적인 업데이트와 최신 기술 정보 제공

Powered By

deepseek · Llama 3 From Meta · Stable Diffusion · LangChain · OpenAI Whisper · LLM · Hugging Face · deepspeed

13

SAI On-Premise / Cloud Deployment 서버 구축



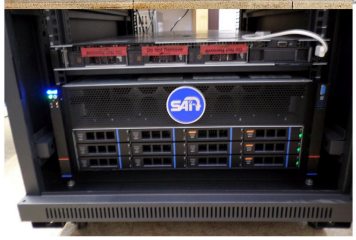


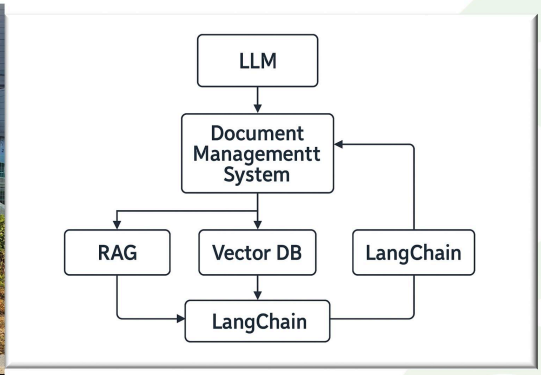


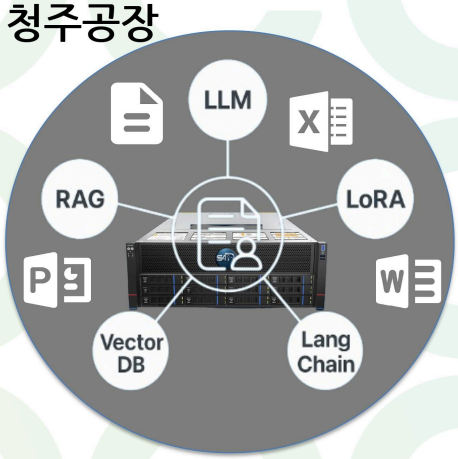
14

SAI On-Premise 서버 구축 사례 삼성 SDI 청주공장









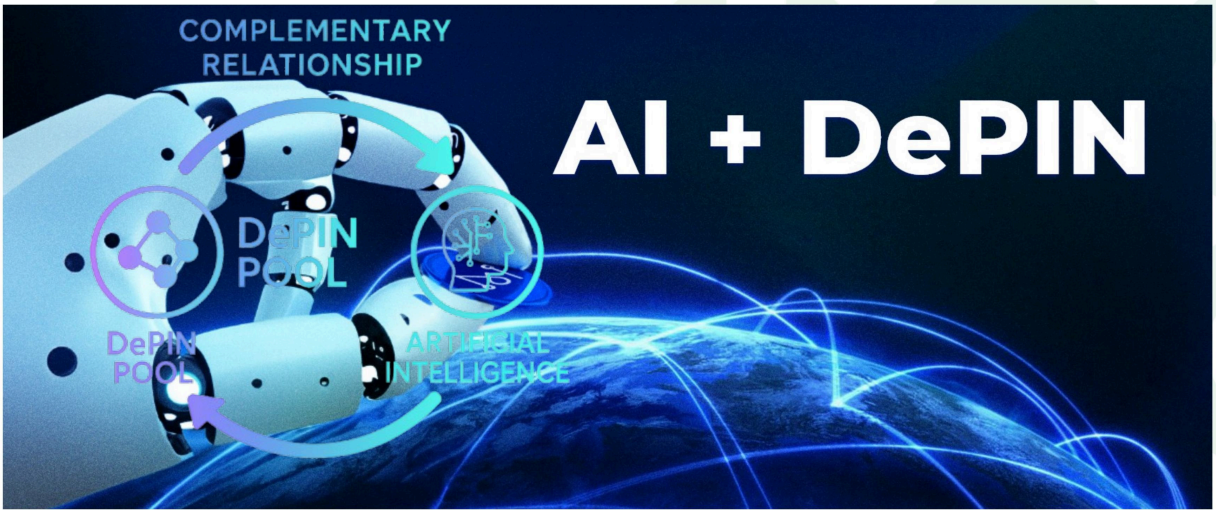
삼성 SDI 청주 공장의 40년간의 생산 데이터를 SAI서버를 통해 RAG와 벡터디비에 저장하여 LLM을 통해 AI 문서 검색 시스템을 성공적으로 구축 하였습니다.

15



AI가 촉발한 DePIN의 수요 폭발

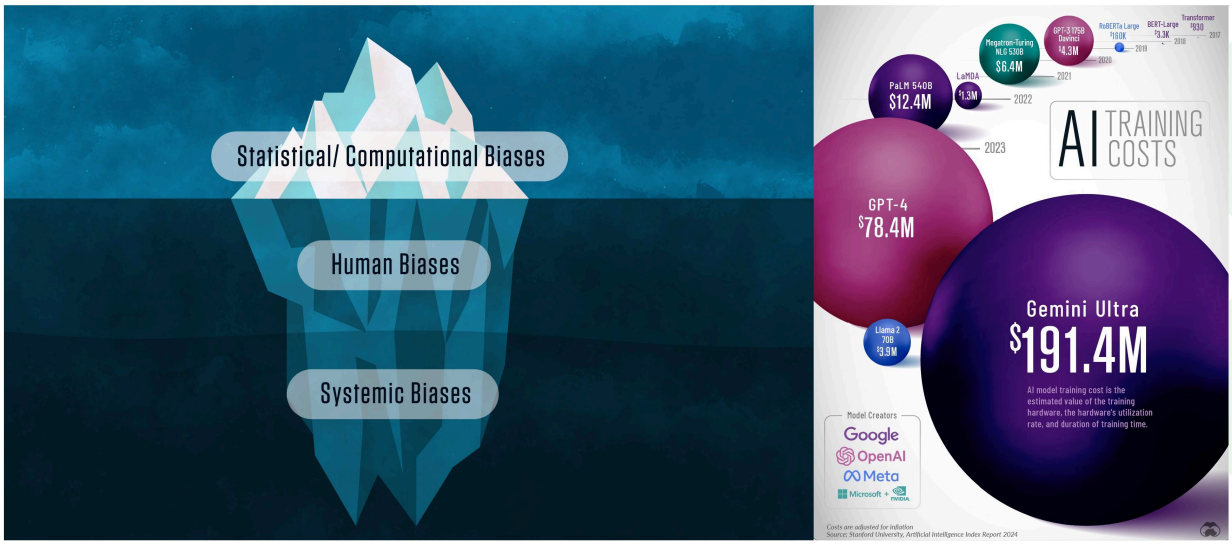
- 인공지능에 필요한 막대한 자원해결은 DePIN이 정답!



17

AI 구축의 가장 큰 고민은? - DePIN으로 해결 가능

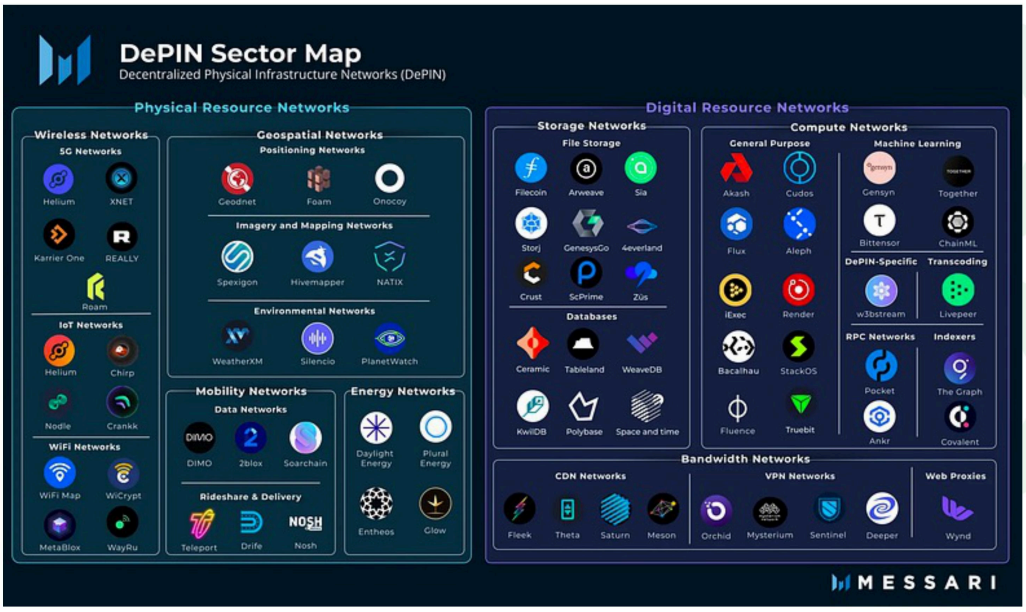
- **학습 편향성** 해결과 과도한 **컴퓨팅 파워 공급비용이** 문제!



18

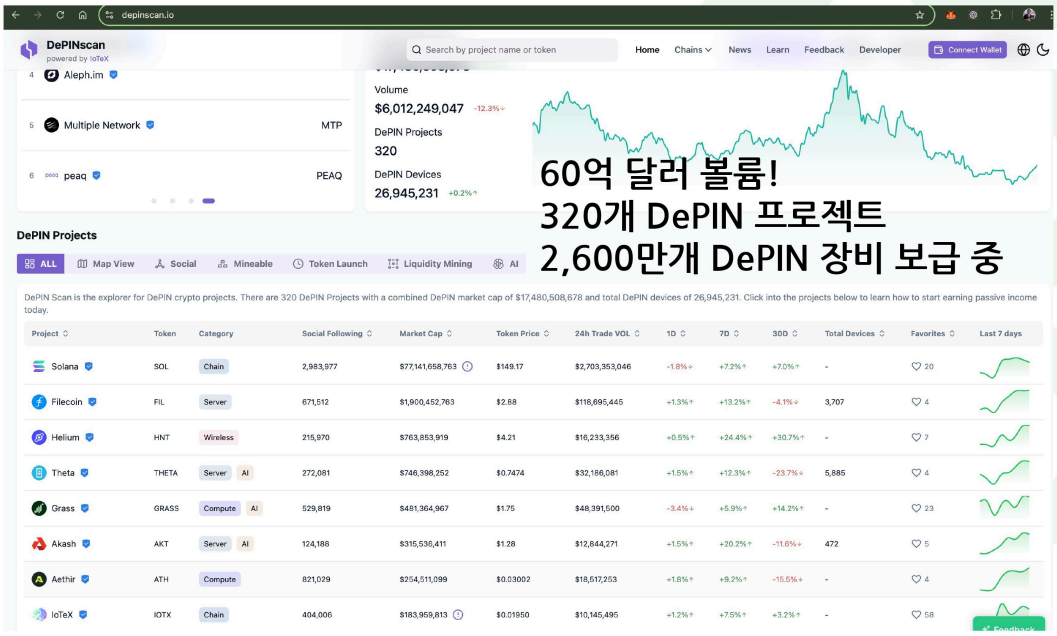
DePIN의 두가지 섹터 분류

- **Physical, Digital** 두가지 형태로 분류 (Messari Report)



19

현제, 320가지 주요 DePIN 프로젝트 가동중!

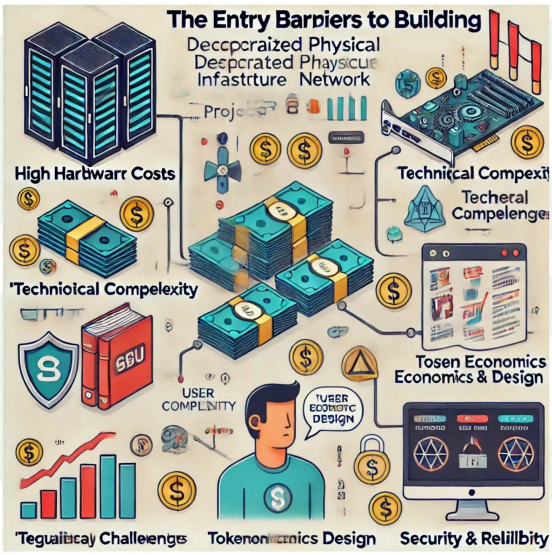


20

너무 높은 DePIN 진입장벽

- 기술과 자본, 관리가 어려워 엄두가 안나는 DePIN 산업

1. **High Hardware Costs** - 서버와 GPU 장비, 고비용문제
2. **Technical Complexity** - 복잡한 기술 설계, 개발 필요
3. **Regulatory Challenges** - 법률 및 규제 문제
4. **User Adoption** - 사용자 이해 부족, 낮은 참여율
5. **Token Economics Design** - 토크노믹스 설계의 어려움
6. **Security & Reliability** - 보안 및 시스템 신뢰성 확보



21

DePIN 프로젝트의 검증 문제

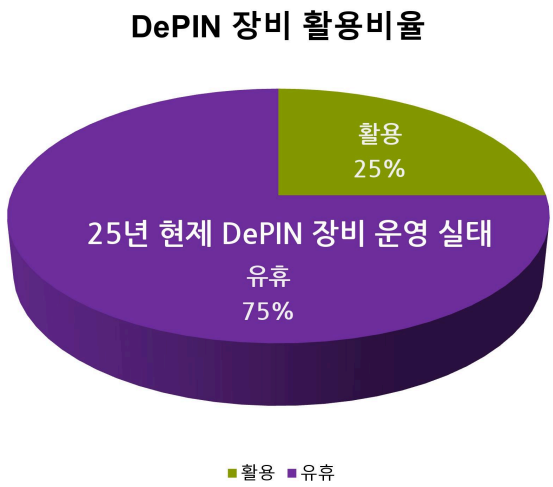
- DePIN 프로젝트라 정의할 수 있는 기준은 무엇인가?

1. **제공 자원 종류** - 서버, 저장장치 등 다양한 리소스
2. **인센티브 메커니즘 검증** - 보상 체계의 설계와 검증
3. **페널티 메커니즘 확인** - 잘못된 행동에 대한 제재 체계
4. **합의 메커니즘 존재 여부** - POW/POS 블록 생성 방식
5. **유지보수 주체** - 누가 네트워크를 관리하는가
6. **토큰 상장 내역** - 주요 거래소 상장 여부 및 유동성
7. **커뮤니티 건전성** - 활발하고 신뢰할 수 있는 사용자 기반



22

전체 배포장비 대비 실제 가동률은? 20~30%?



1. 연산용 DePIN 사례 (OCP 서버)

•DePIN에 참여한 OCP 서버 중 4~12대가 평균 6개월 이상 가동되지 않은 상태 medium.com.

Ex) 기업이 4대 서버를 투입했을 때, **절반 이상(2~3대)이** 장기 유휴 상태로 남는 경우도 있음

2. 네트워크형 DePIN(와이파이·LoRa·센서 등)

•Helium, Grass 등 무선 인프라 DePIN 프로젝트의 경우:

- 수백만 대 장비가 전 세계에 배포되었지만, 실제 **일일 Active 노드**는 그 중 일부만 활성화됨
  - Nosana: 전체 4,200노드 중 **일일 활성 약 600~800노드** followin.io+3research.grayscale.com+3solanafloor.com+3mexc.com.
  - Render Network: 3,784 마이너 중 **수백만 정도가** 실제 주간 또는 일간 사용 상태.

배포된 장비 대비 실제 운영되는 비율은 20~30% 수준

23




ISACA
Korea Chapter
ZETA CUBE
The Solution Leader of DePIN & AI

Track 1-3

불확실성 시대에서의 AI 거버넌스와 리스크관리 전략, 새로운 AI 감사 자격 체계 AAIA를 중심으로



류종기
EY한영 상무

발표개요

인공지능(AI) 기술이 기업 경쟁력을 좌우하는 핵심 동력이 되었지만, 동시에 모델 편향성, 투명성 부족, 윤리 문제 등 새로운 차원의 경영 리스크와 불확실성을 증폭시키고 있습니다. 본 세션은 AI 도입 및 운영의 전 과정에서 발생할 수 있는 주요 리스크를 선제적으로 식별하고 통제하기 위한 **AI 거버넌스 및 리스크 관리 전략**을 심층적으로 다룹니다.

이력

25년간 IBM, Deloitte, EY에서 컨설턴트로 근무, 서강대/울산과학기술원(UNIST)/경기대 겸임교수

AI 거버넌스와 리스크관리 전략

- 새로운 AI 감사 자격 체계 AAIA를 중심으로

류종기
EY한영 | 상무, 한국정보시스템감사통제협회 이사

Confidential. For internal use only.

ISACA Korea Chapter

Agenda

Domain 1: AI Governance and Risk	Domain 2: AI Operations	Domain 3: AI Auditing Tools and Techniques
A. AI Models, Considerations, and Requirements	A. Data Management*	A. Audit Planning and Design
B. AI Governance and Program Management	B. AI Solution Development Methodologies and Life Cycle	B. Audit Testing and Sampling Methodologies
C. AI Risk Management	C. Change Management*	C. Audit Evidence Collection Techniques
D. Privacy and Data Governance Programs	D. Supervision of AI Solutions	D. Audit Data Quality and Data Analytics
E. Leading Practices, Ethics, Regulations, and Standards for AI	E. Testing Techniques for AI Solutions	E. AI Audit Outputs and Reports
	F. Threats and Vulnerabilities*	
	G. Incident Response Management *	
	*Specific to AI	

출처 : ISACA Advanced in AI Audit™ (AAIA™)

2


ISACA
AAIA
ISACA Advanced in AI Audit™
AI Governance and Risk
DOMAIN 1


AI Models, Considerations, and Requirements
Section A

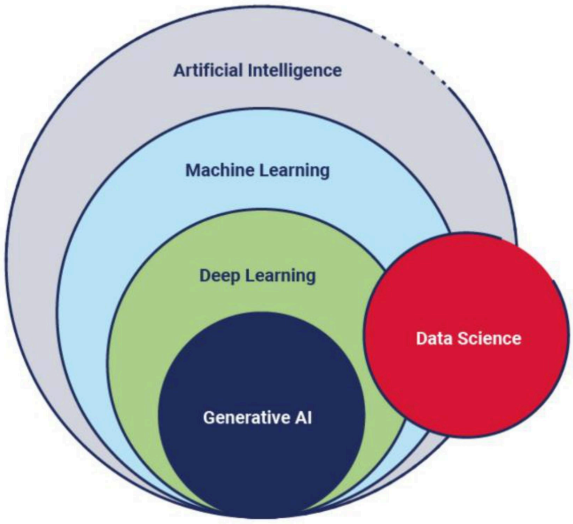
Overview of AI and Related Technologies

Artificial Intelligence (AI)
AI involves techniques that equip computers to emulate human behavior, enabling them to learn, make decisions, recognize patterns, and solve complex problems.

Machine Learning (ML)
ML is a subset of AI and uses advanced algorithms to detect patterns in large datasets, allowing machines to learn and adapt. ML algorithms use supervised or unsupervised learning methods.

Deep Learning (DL)
DL is a subset of ML that uses neural networks for in-depth data processing and analytical tasks. DL leverages multiple layers of artificial neural networks to extra high-level features from raw input data, simulating the way human brains perceive and understand the world.

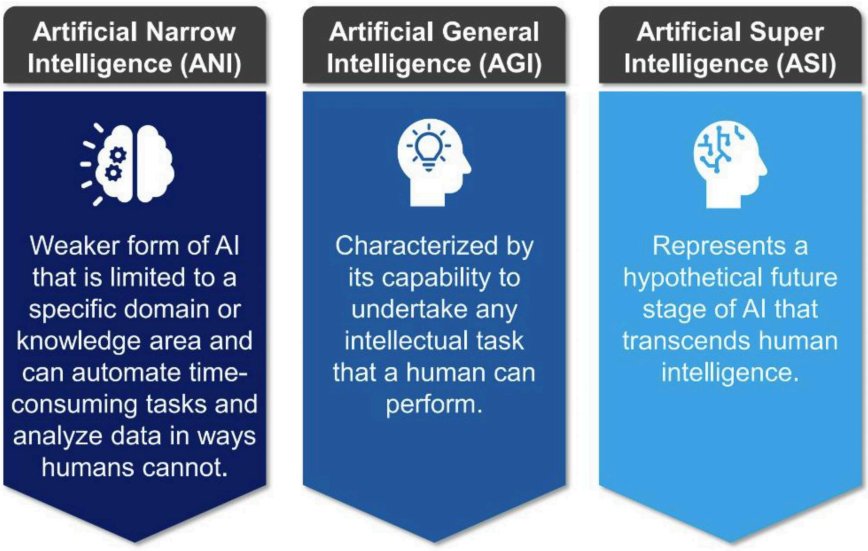
Generative AI
Generative AI is a subset of DL models that generates content like text, images, or code based on provided input. Trained on vast datasets, these models detect patterns and create outputs without explicit instructions, using a mix of supervised and unsupervised learning



출처 : ISACA Advanced in AI Audit™ (AAIA™)

5

Types of AI Based on Capabilities



출처 : ISACA Advanced in AI Audit™ (AAIA™)

6

Predictive Models

- Use historical data and ML algorithms to forecast future events and trends to identify patterns and relationships, and make informed predictions
- Have a wide range of applications (e.g., predictive maintenance in manufacturing, customer churn prediction, and financial forecasting)

Unique Risk of Predictive Models

- Inheriting biases from training data.
- Mishandling data that contains sensitive information.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

7

Generative Models

Can generate new and original content (images, code, text, speech, translations, music, art, etc.)
Builds on patterns and examples from its existing training data to generate new content
Presents its own set of legal and ethical risk:

- Potential distribution of harmful generated content
- Copyright ownership and infringement
- Intellectual property concerns



Both networks' performances continuously improve as training progresses until the generator can produce new data so realistic that the discriminator cannot reliably differentiate it from real data

출처 : ISACA Advanced in AI Audit™ (AAIA™)

8

ML/AI Models

Innovative subfield of AI that:

- Employs models and algorithms to enable machines to learn from data without explicit programming
- Is based on various mathematical concepts (statistics, probability, linear algebra, calculus, and optimization)
- Leverages advanced statistical methods to develop robust algorithms capable of solving complex problems
- Extracts valuable insights from large and diverse datasets used in applications (pattern recognition, classification, anomaly detection, and predictive analytics)

Understanding the inner workings of AI algorithms improves design and implementation of AI to:

- Better solve specific problems
- Achieve desired outcomes
- Help users identify potential limitations or biases
- Take steps to address identified limitations or biases

출처 : ISACA Advanced in AI Audit™ (AAIA™)

9

Neural Networks: Types and Examples

Feedforward neural networks (FNN):
Information flows in one direction (input to output) with no loops or feedback (email filtering, text/image recognition)

Recurrent neural networks (RNN):
Designed for sequential data and have loops allowing information to pass through sequences of steps (language translation, speech recognition, and text generation)

Convolutional neural networks (CNN):
Specialized for processing grid-like data and use convolutional layers to detect patterns (photo tagging, object detection, and medical imaging)

출처 : ISACA Advanced in AI Audit™ (AAIA™)

10

Relationship Between Algorithms and ML Models

Algorithm: A set of step-by-step instructions or rules to be followed to solve a specific problem or perform a particular task:

- Serve as the building blocks that enable models to analyze data, identify patterns, and make predictions
- Have associated settings (hyperparameters) that can be adjusted prior to training the model to optimize its performance
- Facilitate the ML model's ability to adapt and improve over time

ML Models: Constructed using one or more algorithms:

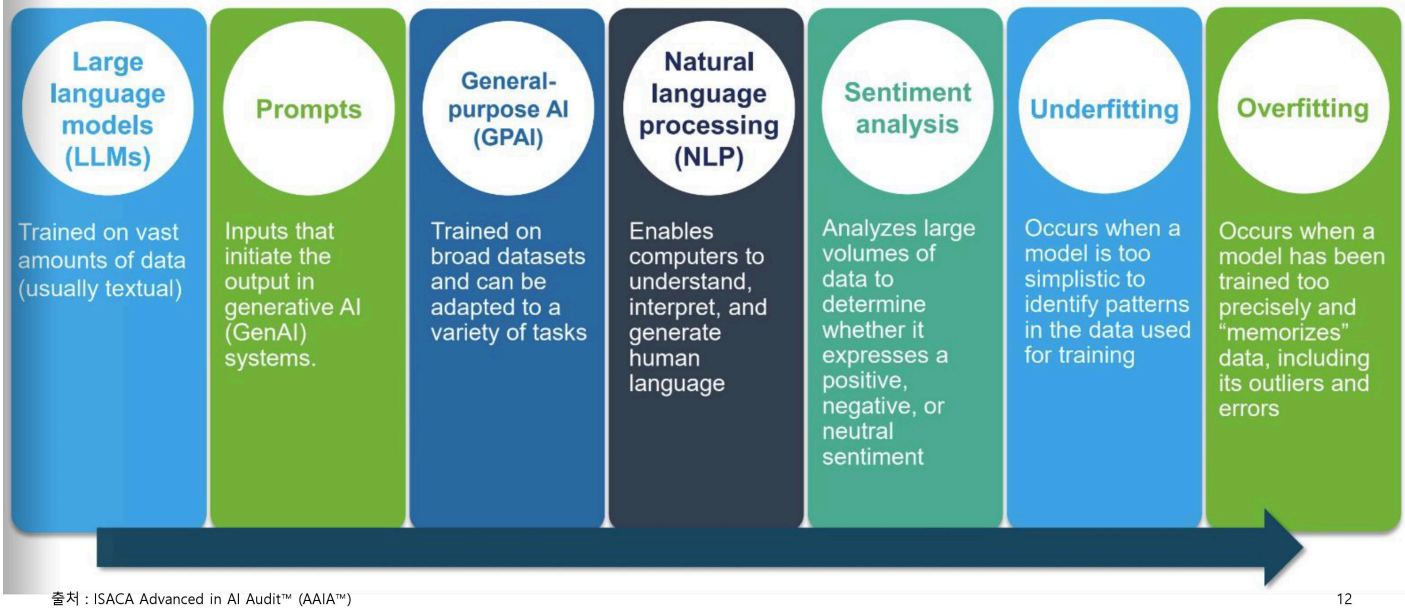
- Learn from and leverage data to make predictions or decisions without explicit programming
- Performance is iteratively refined based on training data

Hyperparameters: Guide the machine learning process:

- Influencing how the algorithm operates (the learning rate or the depth of decision trees in certain models)

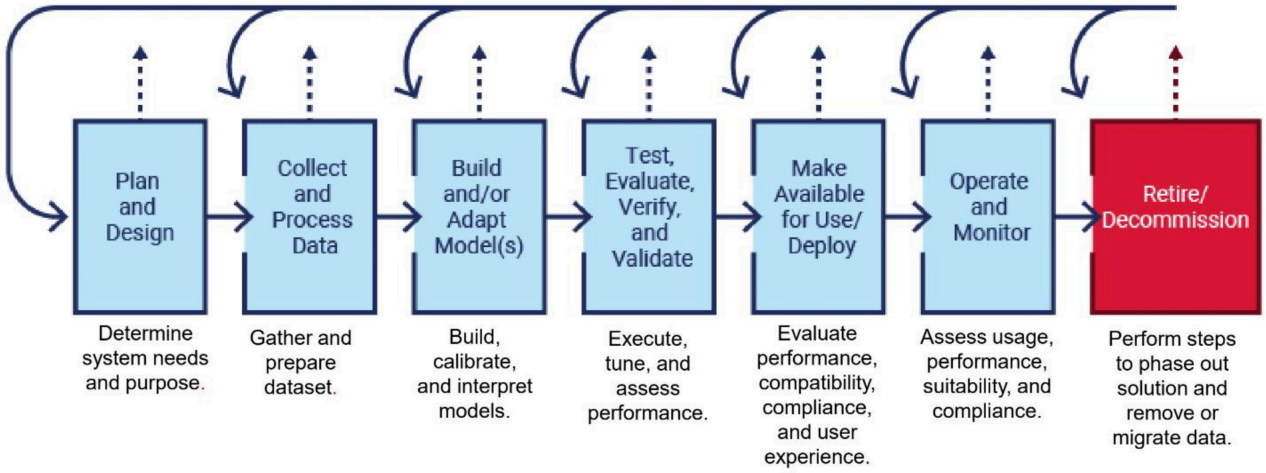
출처 : ISACA Advanced in AI Audit™ (AAIA™) 11

Additional AI Considerations



출처 : ISACA Advanced in AI Audit™ (AAIA™) 12

AI Life Cycle Overview



출처 : ISACA Advanced in AI Audit™ (AAIA™)

13

AI Life Cycle Considerations

In each phase:

- Align choices with business/use cases, organizational needs, and governance requirements
- Consider potential risk and address each identified risk promptly
- Manage changes through established organizational change management processes
- Communicate proactively with stakeholders to set expectations, maintain alignment, and build buy-in
- Evaluate the usefulness, value, and compliance of the solution
- Develop documentation explaining key elements and supports implementation, operation, monitoring, and maintenance tasks

출처 : ISACA Advanced in AI Audit™ (AAIA™)

14



출처 : ISACA Advanced in AI Audit™ (AAIA™) 15

AI Governance and Program Management Overview

Refers to the processes, standards, and guardrails that help ensure AI solutions and tools are safe and ethical.

Define	The organization's overall strategy and roles and responsibilities
Establish	Formal policies and procedures
Provide	Policies and procedures to applicable stakeholders through formal training and awareness
Measure	Effectiveness through program metrics
Report	Effectiveness metrics to the appropriate oversight body

출처 : ISACA Advanced in AI Audit™ (AAIA™) 16

AI Strategies and Examples

An AI strategy is an organization's approach to adopting, developing, and governing AI solutions based on a perceived need to provide guidance to applicable stakeholders.

- Developed using agreed-upon best practices that focus on common concepts of AI ethics
- Created using a cross-functional team that provides guidance from across the organization
- Accepted by senior management for the greatest success

National

- Alignment of ethics and country values
- Guiding laws and regulations
- Promoting national security

Industry

- Promoting industry innovation
- Guiding standards and best practices
- Interpretation of AI regulations and guidance

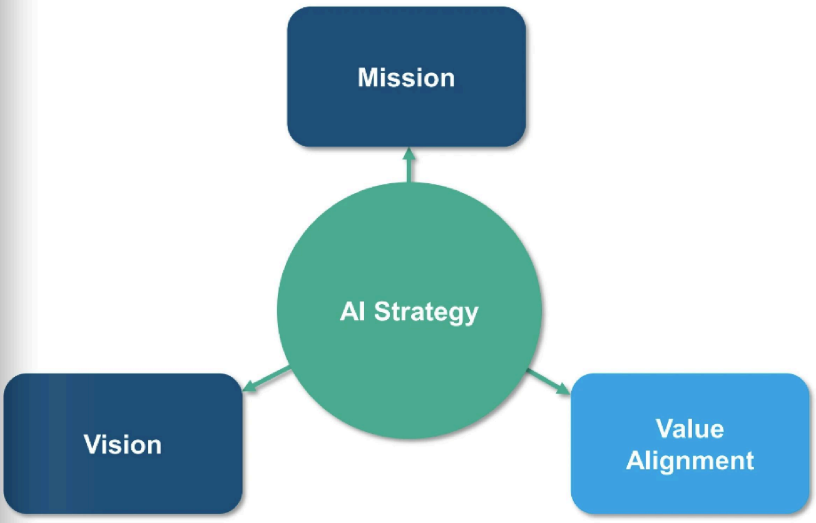
Corporate

- Identifying opportunities for competitive advantage
- Ensuring alignment with applicable requirements

출처 : ISACA Advanced in AI Audit™ (AAIA™)

17

Aligning AI Strategy



Mission:

Defines organization's purpose, goals, culture, and values.

Vision:

Communicates organization's intent to use AI to further their mission.

Value Alignment:

Guides the AI system design to behave in ways consistent with human values and ethical principles. Value alignment should be evaluated throughout the AI life cycle and is achieved through a strong governance process.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

18

AI-related Policies and Procedures

Organizations communicate their expectations surrounding the use of AI to ensure alignment with the AI strategy by:

- **Examining** opportunities to integrate AI concepts into existing policies, standards, and procedures
- **Documenting** a position on AI in a formal policy that is provided to all applicable stakeholders
- **Establishing** authorized and prohibited uses of AI
- **Providing** governance expectations surrounding implemented AI solutions
- **Leveraging** available guidance and generally accepted frameworks
- **Updating** standards, plans, procedures, and employee training with an AI-specific policy rollout
- **Implementing** an AI acceptable use policy (AUP) to guide ethical and responsible deployment of AI.

출처 : ISACA Advanced in AI Audit™ (AAIA™) 19

AI Risk Management

Section C

AI Risk Management

Goal

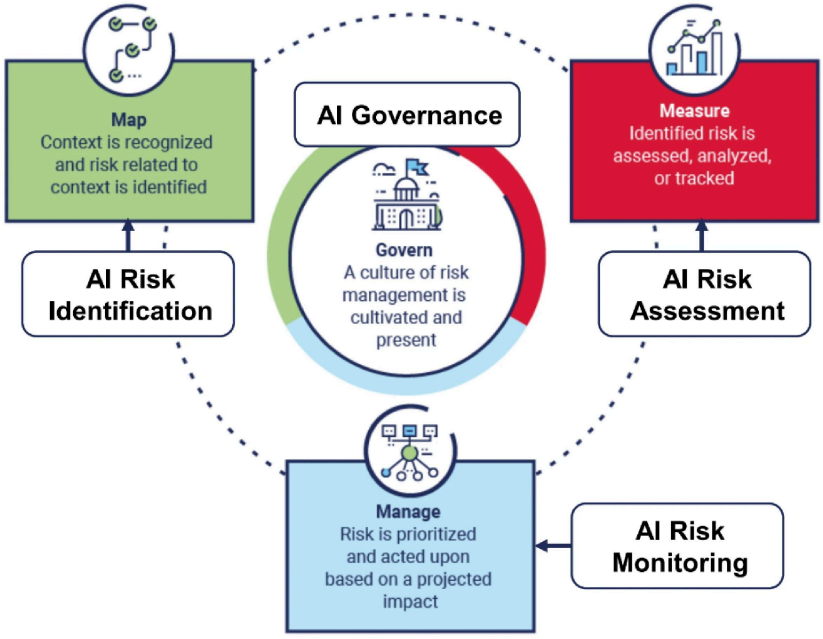
Minimize the potential negative impacts of AI solutions to people, organizations, and the overall ecosystem by proactively addressing, documenting, and managing risk throughout the life cycle.

Method

Use an established framework provided by industry or country.

Example

NIST AI Risk Management Framework Core Functions



Source: US National Institute of Standards and Technology, NIST AI 100-1: Artificial Intelligence Risk Management Framework (AI RMF 1.0), USA, 2023, https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

출처 : ISACA Advanced in AI Audit™ (AAIA™)

21

AI-related Risk Identification

Organizations should remember to:

Integrate
identification into the organization's overall risk management program.

Repeat
The identification process on a regular basis (e.g., onboarding new vendors)

Consider
the unique attributes of AI, including threat actors, risk, and challenges posed.

Employ
an appropriate amount of professional skepticism for any solution being assessed.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

22

AI Threat Landscape

A multifaceted set of risk, challenges, and security/privacy concerns associated with the development and use of AI solutions that involves a wide range of threat actors attempting to exploit vulnerabilities in a new technology.

Examples include:

Insider Threats

Employees causing intentional or unintentional harm (entering confidential information into public AI solutions or extracting sensitive data for financial gain)

Nation States

Used as part of their strategic interest (cyberespionage, warfare, surveillance, censorship)

Cybercriminals

Used to enhance attacks that are more effective or harder to detect (creating convincing phishing emails or developing advanced malware to aid cyberattacks)

AI Developers

Creating solutions that lack considerations for commonly agreed upon ethical standards (a credit rating agency designing a biased or unfair algorithm that unjustly denies credit)

출처 : ISACA Advanced in AI Audit™ (AAIA™)

23

AI Risk

AI poses unique technology and human-related risk to any organization developing or using AI solutions, making risk identification difficult. It is important for organizations to:

- Understand and address risk early in the process to ensure that adverse effects are limited and harm to people, the organization, and the ecosystem are avoided.
- Understand the AI system life cycle as well as the applicable stakeholders present within each phase of that life cycle.
- Use criteria for determining the risk level of an AI application or system (scale, scope, and optionality).
- Consider various criteria and categories of risk for AI systems and their alignment with regulation levels required for AI solutions.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

24

Third-Party Solution Risk Identification

There are multiple risks introduced related to the solution itself and the data it may be using. Examples include:

System

Risk metrics or methodologies used by the AI system developer may not align with risk metrics or methodologies used by the organization

The organization developing the AI system may not be transparent about the risk metrics or methodologies it used.

Data

Risk measurement and management can be complicated by how customers use or integrate third-party data or systems into AI products or services

The organization is at the whim of the third party when it comes to what data was used to train the model.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

25

Risk Assessment

- Perform on a regular basis and during any change to AI solutions
- Leverage existing risk definitions to start assessing the likelihood and impact of identified risk
- Document methods and metrics that measure AI risk and trustworthiness during risk identification
- Evaluate trustworthy characteristics present in the AI solution (ethical use, transparency, explainability, data security, and human rights)
- Identify approaches to track existing, unanticipated, and emerging AI risk
- Gather user feedback about measurement and metric effectiveness

출처 : ISACA Advanced in AI Audit™ (AAIA™)

26

Risk Mitigation and Prioritization

Risk is addressed or treated to an acceptable level based on the organization's risk appetite and tolerance. Options include:

Avoid
Seeks to eliminate risk exposure by not pursuing actions that would incur risk (enterprise chooses not to use AI solutions for product development due to regulatory concerns)

Mitigate
Lessens probability of a risk event or impact using controls (enterprise creates strong AI policies and ensures AI considerations are built into existing security controls)

Share or transfer
Spreads impact of risk across multiple enterprises through partnership or transfers impact of risk through insurance coverage or contractual agreements (enterprise enters into a shared responsibility agreements with a vendor)

Accept
Acknowledges existence of risk and adopts a plan despite possible impact, which can be unavoidable (enterprise management agrees to implement chatbots despite negative customer opinion on AI)

출처 : ISACA Advanced in AI Audit™ (AAIA™)

27

Remediation Plans/Best Practices

Integrate plans with the organization's risk management function and consider all risk applicable to the organization.

For all identified risk that exceeds organizational appetite and tolerance:

Create defined remediation plans with clear steps and expected conclusion data.

Ensure plans are assigned to a risk owner responsible for executing and reporting details to the organizational oversight function.

Consider compensating controls where remediation timelines are longer than acceptable.

Track and provide oversight of plans and assign accountability to an independent organizational function for regular status reporting.

Verify plan completion independently.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

28

Key Risk Indicators (KRIs)

Provide early warnings for potential incidents. This can help an enterprise identify exposure to risk events that may harm business performance and can trigger alerts to relevant personnel.

Purchased:

- Adapts standard setting processes that follow expected business outcomes
- Coordinates with other metrics related to performance (customer satisfaction, resolution rate, misunderstanding rate, escalation rate)

Developed:

- Relies on statistical analysis
- Requires input from cross-functional teams
- Depends on whether solution is white or black box and the model method

출처 : ISACA Advanced in AI Audit™ (AAIA™) 29

Examples of KRIs

Indicator	Description	Target
Customer satisfaction score	Measures overall customer satisfaction with chatbot interactions	95% satisfaction rate or higher
Resolution rate	Percentage of customer inquiries resolved by the chatbot without human intervention	At least 90% resolution rate
Misunderstanding rate	Percentage of interactions where the chatbot fails to understand or respond accurately	Less than 1% misunderstanding
Escalation rate	Percentage of chat sessions escalated to human support due to chatbot limitations	Less than 10% of interactions

출처 : ISACA Advanced in AI Audit™ (AAIA™) 30

Privacy and Data Governance Programs

Section D

Privacy and Data Governance Programs

AI solutions can help streamline data governance and privacy programs within an organization to assist with tasks, including:

Automated data discovery and classification

- Safeguards data if accessed by an unauthorized third-party
- Increases in available computing power allows individuals to more easily attempt to break encryption

Data quality management

- Manufactured audio and video created using digitally manufactured imitations of a person based on samples
- Difficult to distinguish
- Poses challenges to enterprises that require verbal approval

Policy enforcement and compliance monitoring

- Reduced storage costs and increased connectivity led to large expansion in data volume
- Data analysis is necessary for many enterprises
- Enterprises must consider impact of privacy

출처 : ISACA Advanced in AI Audit™ (AAIA™)

32

Analyzing Risk Scenarios

Quick detection and action toward risk is critical to contain the impact:

Examples:

- General Data Protection Regulation (GDPR) – European Union
- California Privacy Rights Act (CPRA) – California, United States
- Personal Information Protection Law (PIPL) – China

Common Themes:

- Entire methodology built on the OpenFAIR model
- Accounts for loss magnitudes at a discrete level
- Factors in control objective maturity as a method to account for potential reductions in overall loss magnitude estimations

출처 : ISACA Advanced in AI Audit™ (AAIA™)

33

Data Privacy Concerns

Privacy risk may arise when developers use large datasets containing sensitive data to train ML models:

Example:

ML models remember data through their accumulated query responses

Result:

Vulnerability to inference where threat actors can determine if sample was contained in the dataset (membership inference) or what sensitive traits a data subject possesses (attribute inference)

Considerations:

Liability for successful inference and conflicting requirements between relevant regulations (GDPR vs. CPRA)

Goal:

Balance potential efficiencies of ML and AI with potential penalties for noncompliance with privacy laws and regulations in their jurisdiction

Action:

Assess the impact of data processing operations on personal data protection

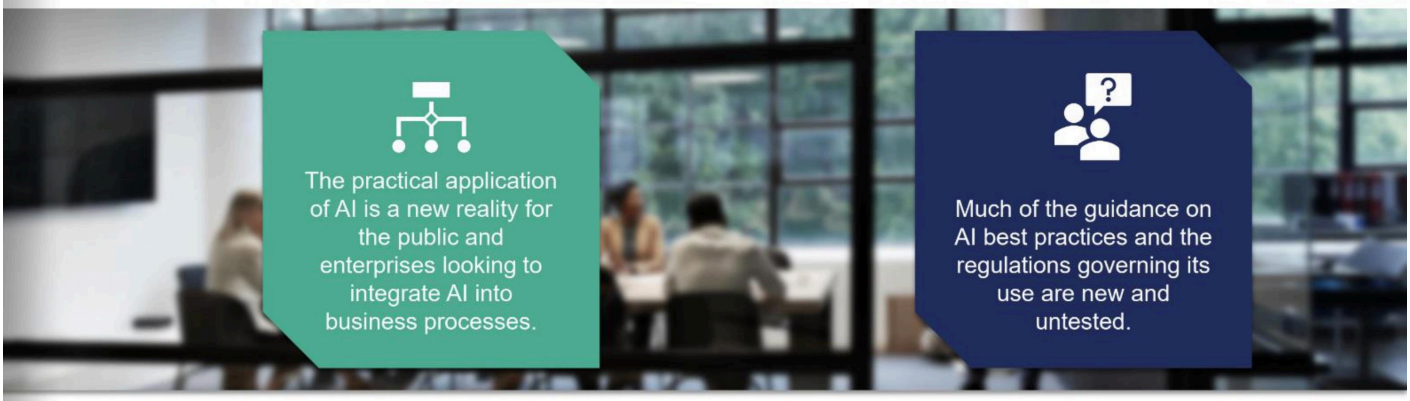
출처 : ISACA Advanced in AI Audit™ (AAIA™)

34

Leading Practices, Ethics, Regulations, and Standards for AI

Section E

Leading Practices, Ethics, Regulations, and Standards for AI



Organization that build or buy AI solutions need to navigate changing regulations, review academic research, implement various standards and frameworks, and consider the potential ethical concerns as these solutions are developed, implemented, and used.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

36

Industry Standards and Frameworks

Designed to assist organizations in their governance, compliance, and operational processes. Examples include:

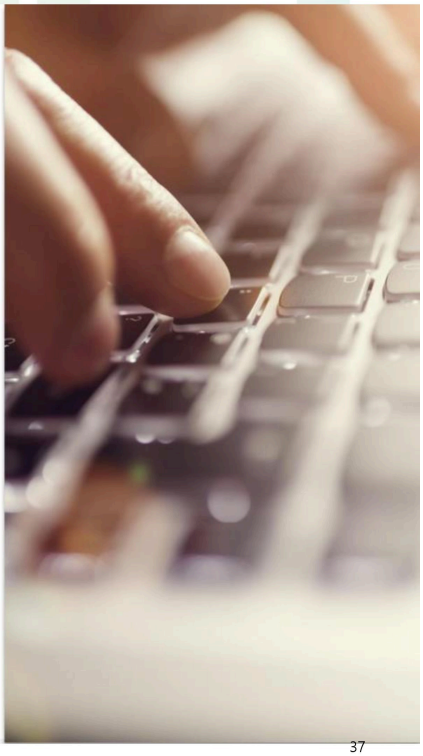
ISO/IEC 23053: Framework for Artificial Intelligence Systems Using Machine Learning: Describes a generic AI system using ML technology

IEEE 7000-2021: Ethical Considerations in AI Systems Design: Focuses on embedding human values into AI, providing AI developers a framework to consider ethical implications during the life cycle of an AI system

OECD AI Principles: Provides a set of recommendations adopted by the membership of the Organisation for Economic Co-operation and Development (OECD) representing various nations

NIST AI Risk Management Framework: Assists organizations in identifying, assessing, and mitigating risk in AI systems

Model AI Governance Framework: Assists organization in deploying AI responsibly



출처 : ISACA Advanced in AI Audit™ (AAIA™)

37

Laws and Regulations

Provide oversight capabilities for governments and their respective enforcement agencies

Many are new and untested or still in draft format

Point to standards and frameworks set by their own intergovernmental agencies to provide compliance guidance

Noncompliance can result in fines or penalties for developers and users

Examples include:

AI Act* (EU):
Categorizes applications based on risk levels

AI Liability* Directive* (EU):
Clarifies civil liabilities for damages caused by AI solutions

Algorithmic Accountability Act* (US):
Requires companies to assess impacts of automated decision systems and mitigate risk

The National AI Act (US):
Aims to enhance AI research, development, and deployment

AI Development Plan (China):
Outlines the country's strategy for AI development, including ethical guidelines and regulatory measures.

*Proposed legislation

출처 : ISACA Advanced in AI Audit™ (AAIA™)

38



Agenda

Domain 1: AI Governance and Risk	Domain 2: AI Operations	Domain 3: AI Auditing Tools and Techniques
A. AI Models, Considerations, and Requirements	A. Data Management*	A. Audit Planning and Design
B. AI Governance and Program Management	B. AI Solution Development Methodologies and Life Cycle	B. Audit Testing and Sampling Methodologies
C. AI Risk Management	C. Change Management*	C. Audit Evidence Collection Techniques
D. Privacy and Data Governance Programs	D. Supervision of AI Solutions	D. Audit Data Quality and Data Analytics
E. Leading Practices, Ethics, Regulations, and Standards for AI	E. Testing Techniques for AI Solutions	E. AI Audit Outputs and Reports
	F. Threats and Vulnerabilities*	
	G. Incident Response Management *	
	*Specific to AI	

출처 : ISACA Advanced in AI Audit™ (AAIA™)

40

Data Management Specific to AI

Section A

출처 : ISACA Advanced in AI Audit™ (AAIA™) 41

Data Management and AI

The foundation of AI begins with good data management hygiene and controls:

- Many issues and risk associated with AI can be linked to issues with data.
- Data plays an essential role in both the development of AI systems and models and its use via data input to elicit a response.
- AI is dependent on data to “teach” the important features and to make decisions.

Relevant areas regarding data include:

- Collection
- Classification
- Confidentiality
- Quality
- Balancing
- Scarcity
- Security

출처 : ISACA Advanced in AI Audit™ (AAIA™) 42

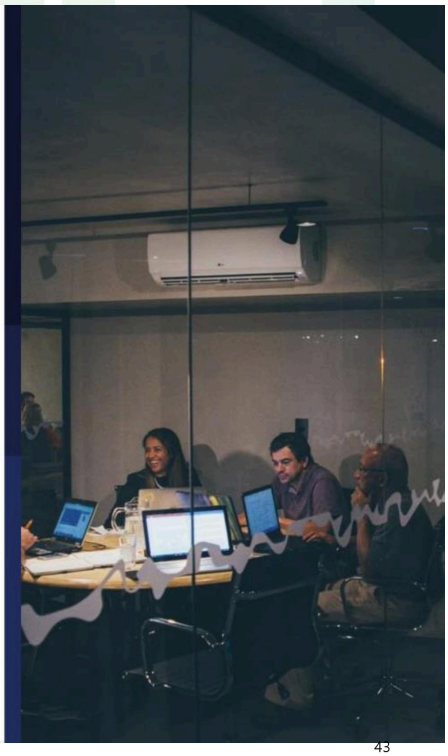
The Five V's of Big Data

Problems arise in how enterprises digitize manual and paper-based data.

Organizations need a mechanism to collect, connect, and centralize data that will be processed and used to train AI models.

Aggregating and collecting data in data warehouses, data lakes, or other solutions helps with data accessibility.

AI models require large volumes of variety and veracity data depending on how the AI model algorithm is being used and developed.



출처 : ISACA Advanced in AI Audit™ (AAIA™)

43

The Five V's of Data (continued)

Examples of AI and data in healthcare:

Velocity	Volume	Value	Variety	Veracity
The speed of data creation and transmission	The amount of available data	What an enterprise can do with the data	The diversity of the data	• Quality of data • Accuracy of data • Integrity of data • Credibility of data
Real-time health monitoring AI requiring constant data updates	AI models that process terabytes of medical imaging data for diagnostics	AI models reducing hospital readmissions by analyzing patient trends	AI that integrates structured patient data (e.g., EHR) and unstructured data (e.g., doctor's notes, images)	Addresses data biases in AI-driven health predictions

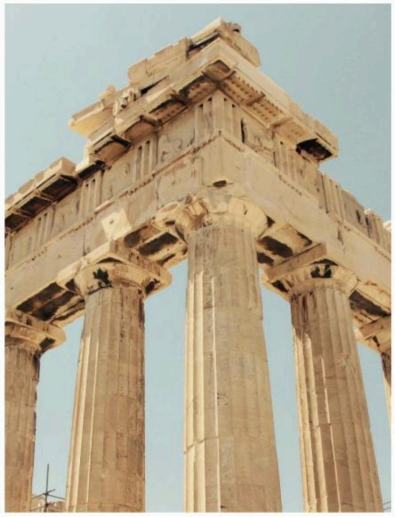
출처 : ISACA Advanced in AI Audit™ (AAIA™)

44

Data Security

Data plays a critical role in the development of AI solutions, but data security takes a more prominent role over the security of the AI solution.

Organizations are required to implement privacy regulations and localized data residency laws have limited the use of personal data.

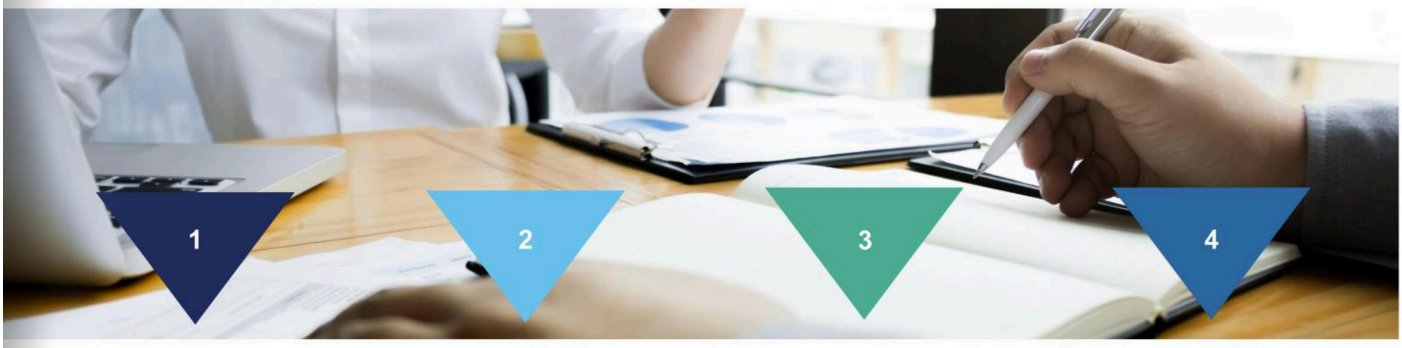


Implementing AI solutions requires additional data security considerations. Given the role that data plays in AI, organizations should evaluate data security risk and controls throughout the AI life cycle.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

45

Data Access



Data is collected from various data sources, which often have their own access and encryption controls in place.

The organization should review the AI system's data flow to identify systems, data stores, and data users.

Additional access controls may need to be applied to systems that are part of the data flow to ensure the congruency of the access controls from originating systems to target systems.

Networking access policies should be reviewed in parallel to ensure containment of the data only to known and authorized systems.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

46

Compromised Data Integrity Scenarios

Data poisoning

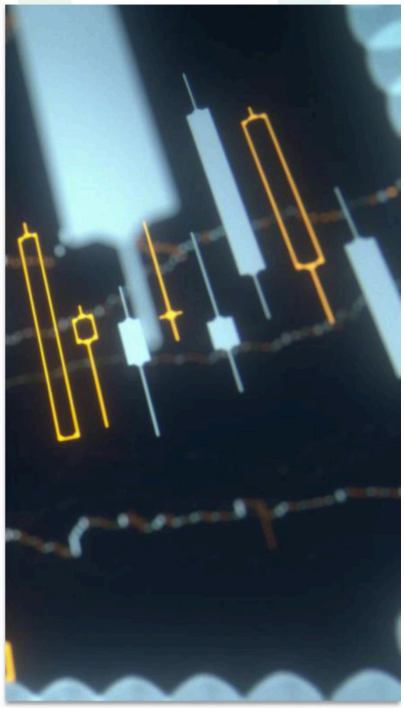
An attacker modifies the training data to influence the model's behavior.

Model tampering

The model weights, architecture, and parameters are modified to provide incorrect, biased, or malicious results.

Embedding tampering

The embedding matrices are tampered with to skew the results of the model.



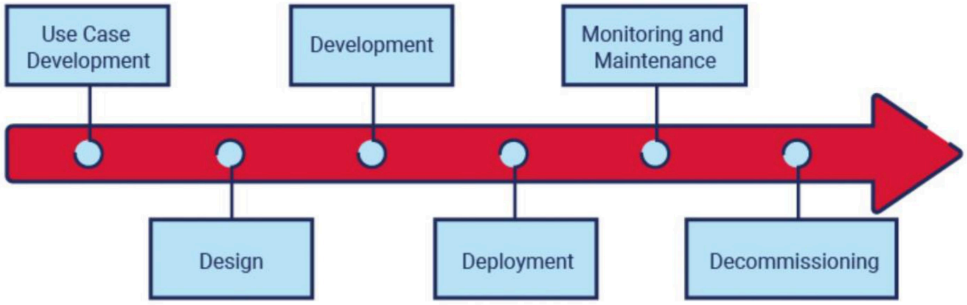
출처 : ISACA Advanced in AI Audit™ (AAIA™) 47

AI Solution Development Methodologies and Life Cycle

Section B

출처 : ISACA Advanced in AI Audit™ (AAIA™) 48

AI Solution Development Life Cycle

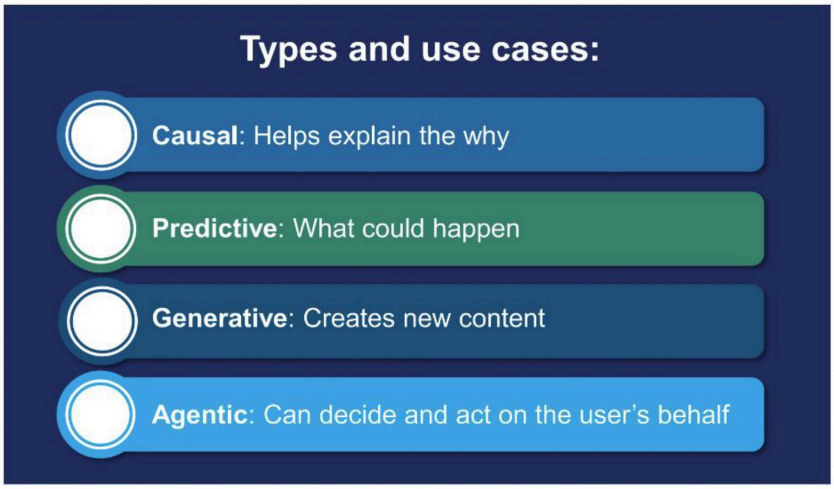


출처 : ISACA Advanced in AI Audit™ (AAIA™)

49

Use Case Development
AI Solution Development Life Cycle

Good use cases drive return on investment (ROI) and are aligned with value for the organization and success factors used to solve business problems.



출처 : ISACA Advanced in AI Audit™ (AAIA™)

50

Good AI Model Use Cases Attributes

Problem-focused	Data Quality	Business and Technical partnership
• Address business problem to solve vs. technology • Commit AI project resources with adequate time to solve the problem • Understand limitations of AI	• Source data of sufficient volume and veracity relevant to the use case • Prepare cleaned, organized, and properly labeled data	• Assign technical resources who understand the business intent and domain context of the problem • Ensure business resources effectively communicate the business problem and success metrics (not AI model performance)

출처 : ISACA Advanced in AI Audit™ (AAIA™)

51

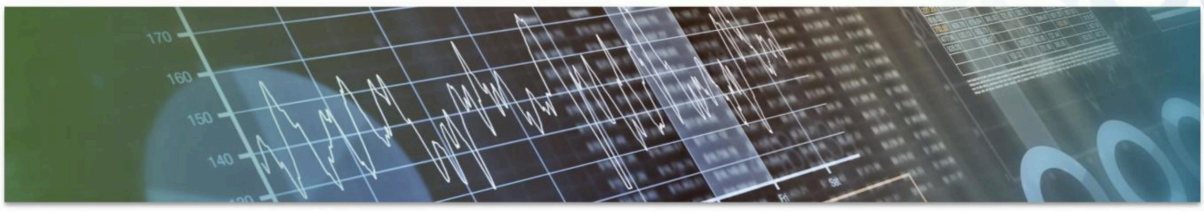
Design
AI Solution Development Life Cycle

Involves

- Exploration of dataset to determine suitability to address business case and solve the identified problem through iterative stages
- Collaboration between the business and data subject matter experts who understand how to apply mathematical algorithms to solve problems

Data Ingestion: Aggregating data from multiple sources into centralized data store for easier exploration and correlation

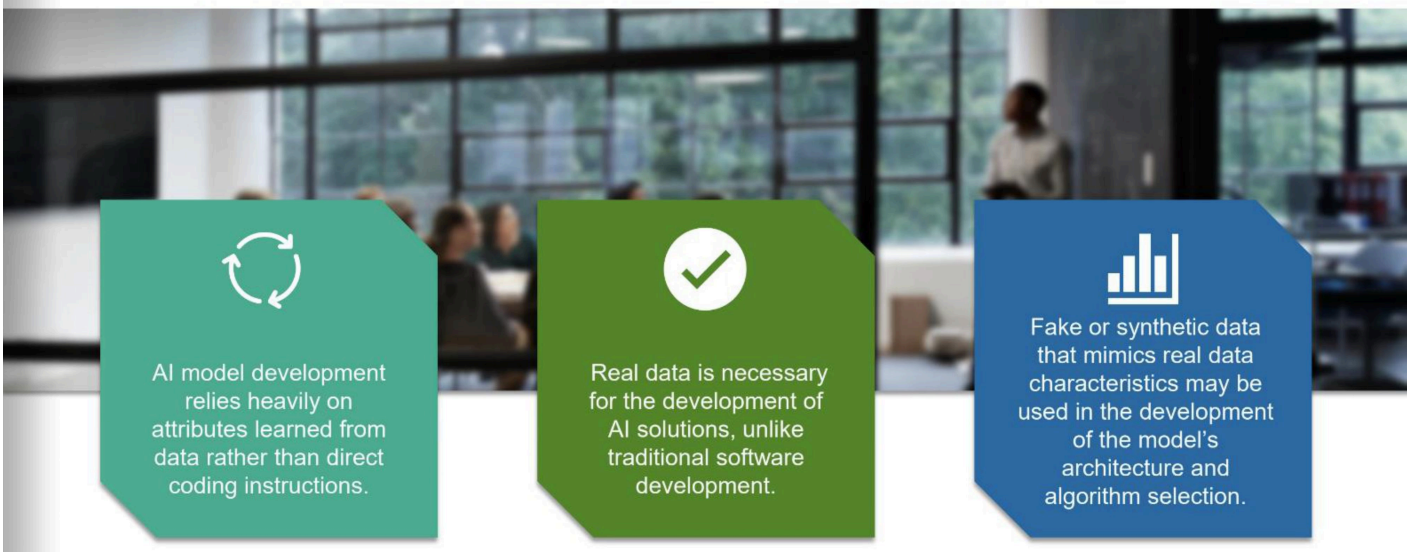
Feature Selection: Leveraging one or more models, like ML algorithms per the use case, to recognize and learn patterns from provided datasets



출처 : ISACA Advanced in AI Audit™ (AAIA™)

52

Development



AI model development relies heavily on attributes learned from data rather than direct coding instructions.

Real data is necessary for the development of AI solutions, unlike traditional software development.

Fake or synthetic data that mimics real data characteristics may be used in the development of the model's architecture and algorithm selection.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

53

Data Preprocessing

Often performed to clean and resolve data quality issues, including:

- Missing values
- Normalizing data
- Removing duplicates or outliers to standardize data

Additional preprocessing steps should be included to ensure that the training of AI models is ethically and responsibly developed.

Some deep learning models, especially LLMs, require the raw dataset to be tokenized.

Dataset Preparation

- Involves the collection of a labeled dataset with relevant and diverse examples that align with the use case objectives
- Preparing high-quality and diverse data is essential to the performance of the model in real-world applications
- Splitting the dataset into a minimum of three chunks is necessary to maintain the integrity of the training and evaluation phases of AI development

출처 : ISACA Advanced in AI Audit™ (AAIA™)

54

Evaluation

Validation Dataset

- Evaluate model performance.
- Test recall and precision.
- Use new data to determine if model performance is sustained and generalizes well against unseen data.
- Perform several iterations of training, tuning, and retraining to attain suitable performance metrics.

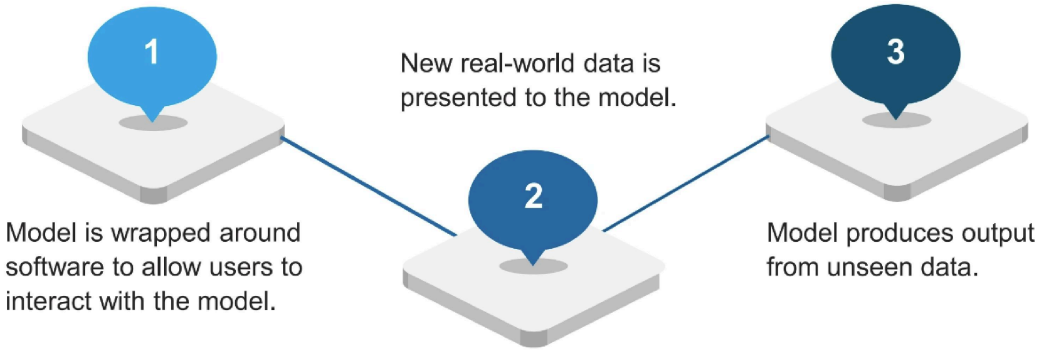
Testing Dataset

- Evaluate final performance.
- Evaluate model for harmful and malicious outputs.
- Maintain artifacts from testing and evaluation to support explainability.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

55

Deployment

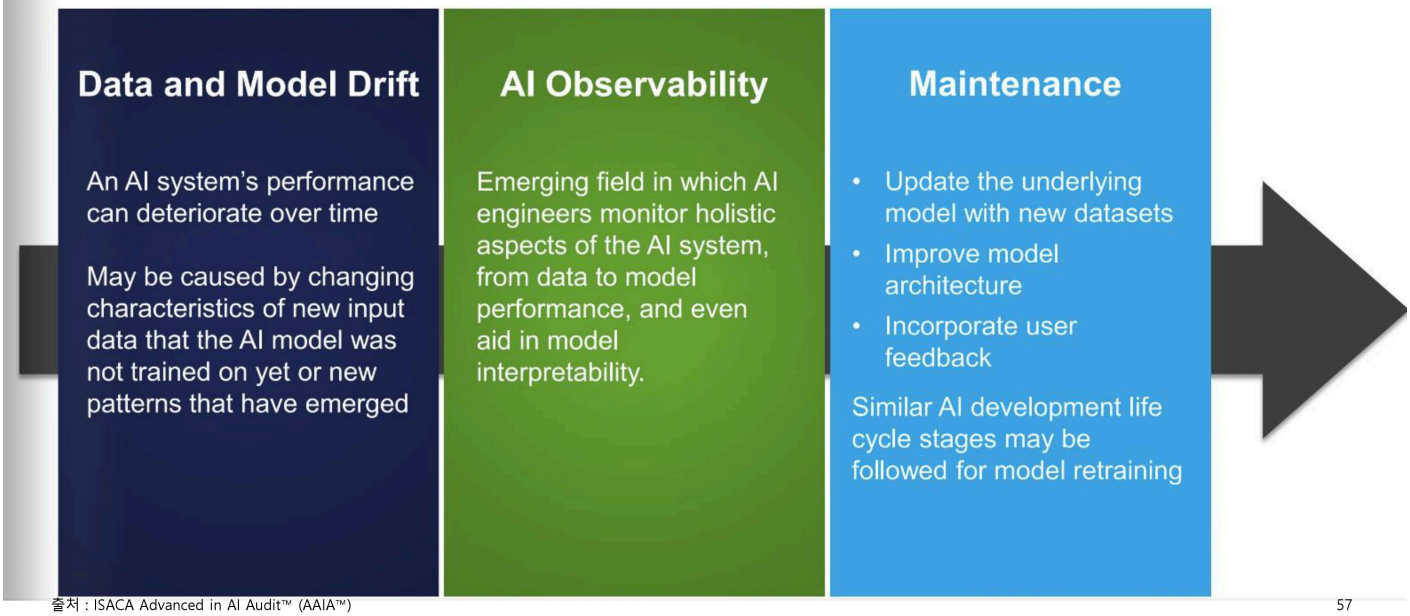


Additional data processing may be needed depending on the AI user's input modality (text, speech, visual).

출처 : ISACA Advanced in AI Audit™ (AAIA™)

56

Monitoring and Maintenance



출처 : ISACA Advanced in AI Audit™ (AAIA™)

57

Decommission

Ensure derivatives of the base AI model are considered in scope for program decommissioning.



Parts of the AI model could be "forgotten" through training without specific datasets, but this can be impractical for larger models.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

58

Change Management Specific to AI

Section C

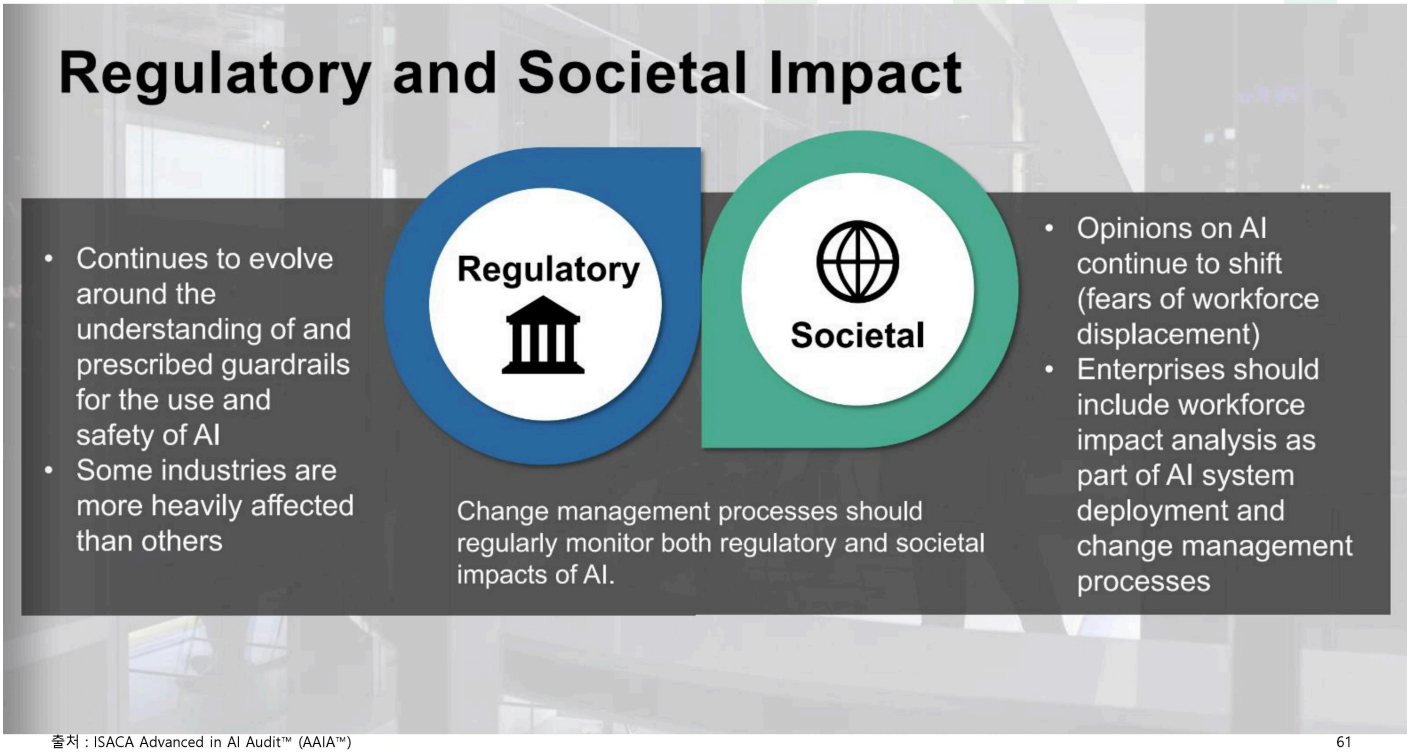
AI Model

AI system behavior can be altered by changing the underlying model(s).

Models can be changed as new model versions are trained and developed.		Governance of the underlying AI model should be part of any organizational change management program.
	As AI researchers develop and train new model versions, models can be swapped in and out which can dramatically alter its output.	

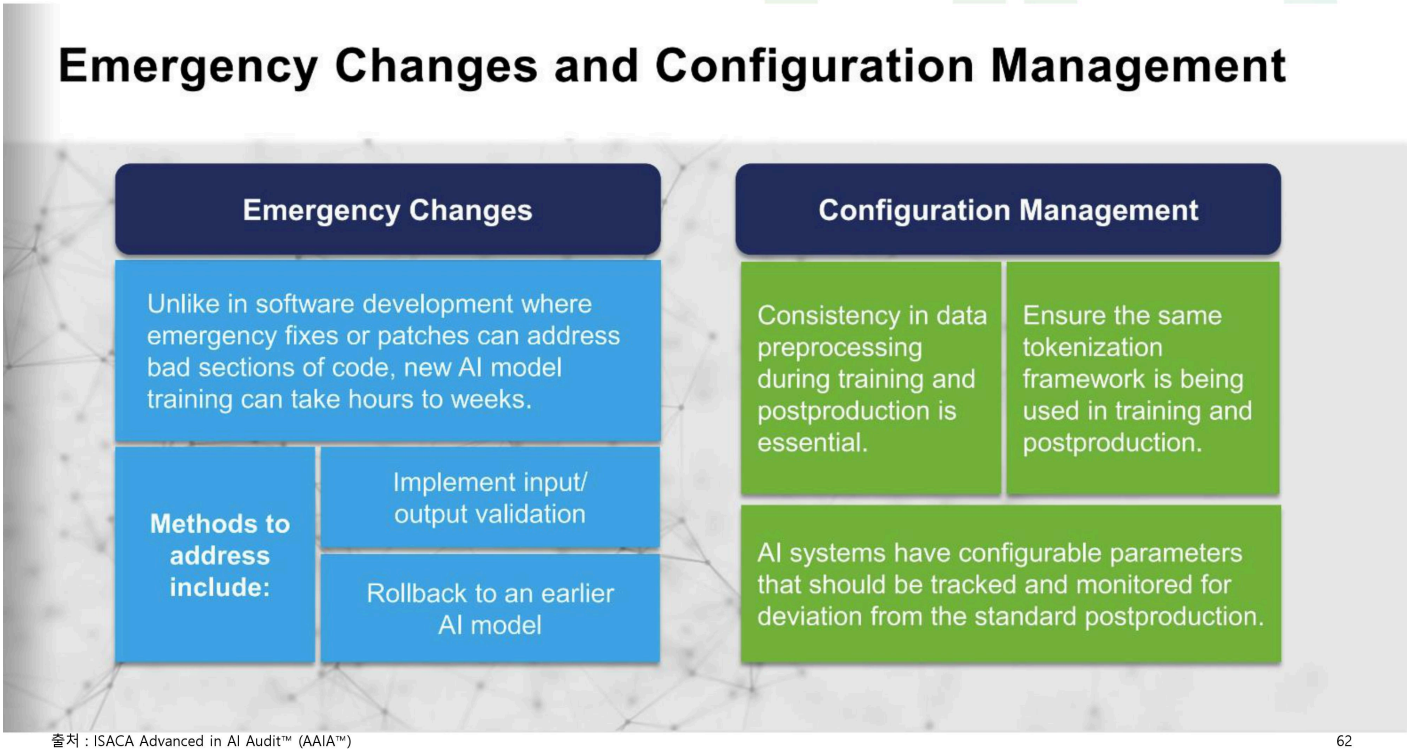
출처 : ISACA Advanced in AI Audit™ (AAIA™)

60



출처 : ISACA Advanced in AI Audit™ (AAIA™)

61



출처 : ISACA Advanced in AI Audit™ (AAIA™)

62

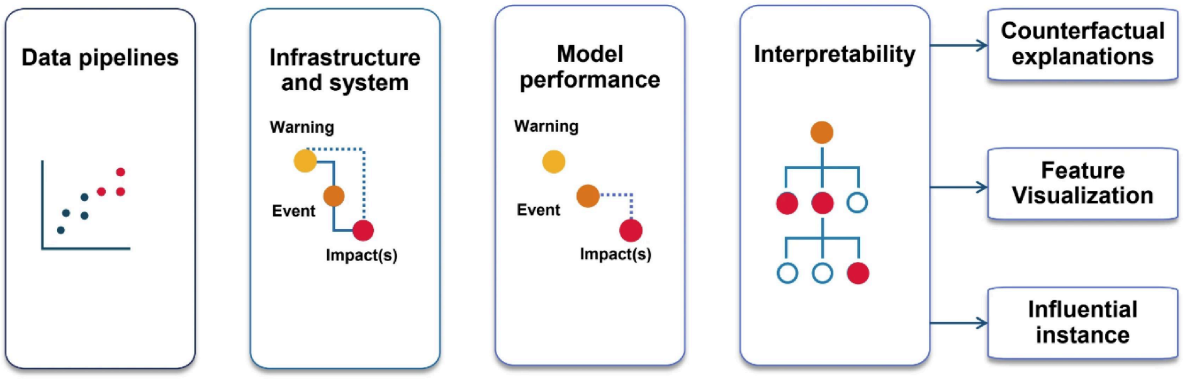
Supervision of AI Solutions

Section D

AI Observability

Helps ensure the availability, reliability, performance, and trustworthiness of the AI system.

Categories of observability include:



출처 : ISACA Advanced in AI Audit™ (AAIA™)

64

Humans in the Loop (HITL)

HITL is a control that puts a human in the final decision-making role to approve of the AI workflow.

This control is designed and implemented in the AI system to require a HITL for specified conditions and explicit approval before the workflow is allowed to continue

The inclusion human oversight helps mitigate risk and ensures that the AI's actions are aligned with human judgment in high-risk situations.

Example of a HITL for an autonomous vehicle:

01

AI Decision: The AI system detects an unexpected obstacle (e.g., pedestrians).

02

Human Approval: AI system sends an alert to the driver, for approval of emergency stop.

03

Execution: If approved, the vehicle performs the emergency stop; if not, the system must adjust its actions accordingly.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

65

Hallucination

A hallucination is when a model makes a prediction that is not grounded in the input data or reality. This can be caused by:

- Training data
- Ambiguity in data input
- Lack of real-world grounding

Controls include:

- Using good quality, diverse, and relevant data
- Implementing guardrails for input and output
- Using templates for output
- Implementing strict data validation protocols during data preprocessing
- Making the model's decision-making process more transparent

출처 : ISACA Advanced in AI Audit™ (AAIA™)

66

Testing Techniques for AI Solutions

Section E

AI Testing Techniques

Model Cards: Provide transparency about model details, including architecture, training datasets, performance metrics, and use case limitations. These are essential for developing targeted testing and ensuring fair AI solutions.

Bias Testing: Detects and corrects biases inherited from training data that can lead to hallucinations, inaccuracies, or harmful outputs. Biases should be identified and mitigated early in the data preprocessing and cleansing stages.

Adversarial Testing: Involves testing the AI system by intentionally feeding it malicious or unexpected inputs to assess its resilience and identify vulnerabilities in edge cases.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

68

Threats and Vulnerabilities Specific to AI

Section F

Types of AI-related Threats

AI threats fall into three major attack surfaces:

Development-time threats

- Occur during the development of the AI system
- Attack surfaces include the engineering environment and the supply chain

Runtime security threats

- Attacks that occur due to non-AI-specific security weaknesses
- Remember that AI systems are also IT systems
- Consider common IT vulnerabilities associated with software wrappers or other IT infrastructure hosting the AI solution

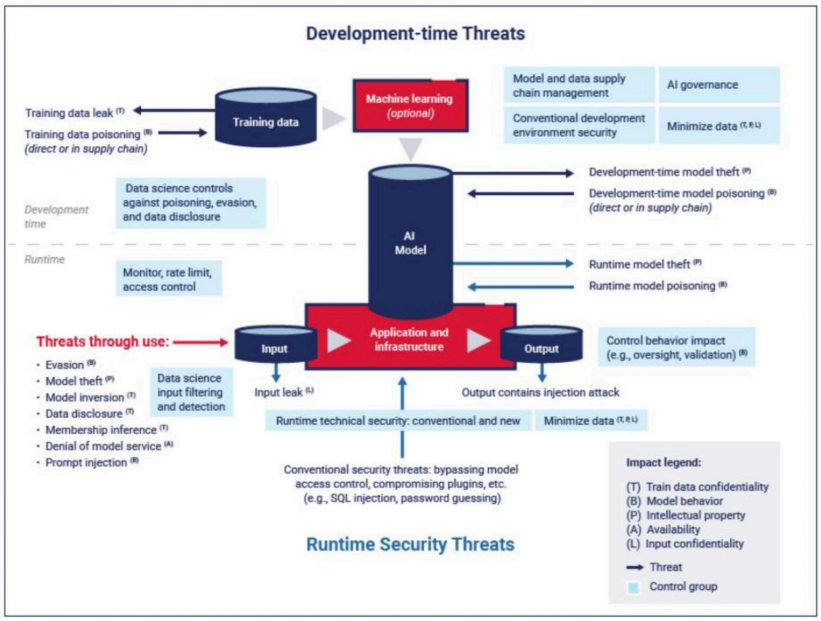
Threats through use

- Occur during normal activities associated with the use of an AI model
- Input and output activities are of particular concern

출처 : ISACA Advanced in AI Audit™ (AAIA™)

70

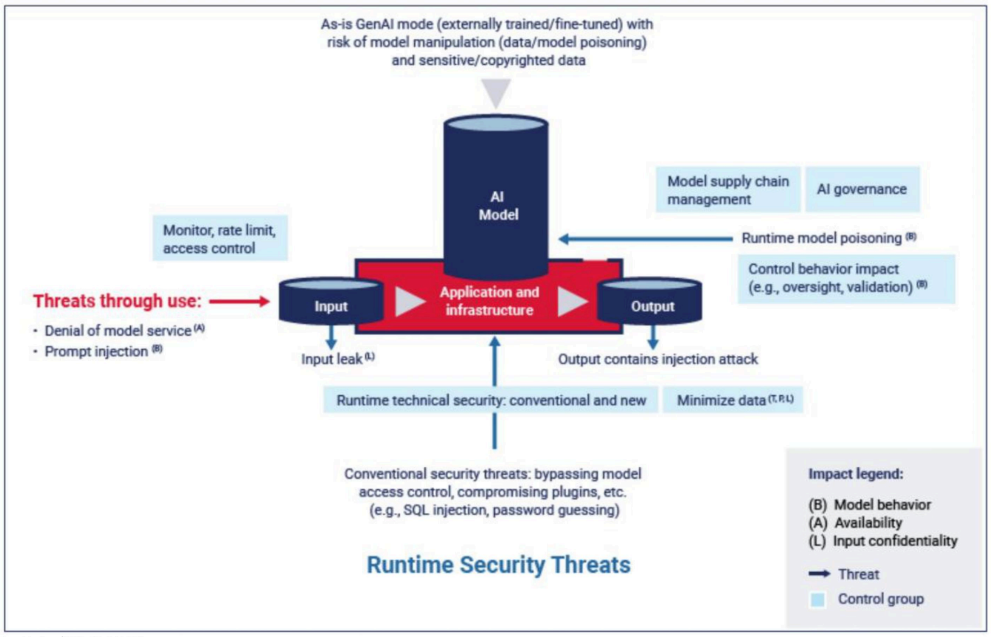
Types of AI-related Threats



출처 : ISACA Advanced in AI Audit™ (AAIA™)

71

Threats for Using Vendor-Supplied AI



출처 : ISACA Advanced in AI Audit™ (AAIA™)

72

Incident Response Management Specific to AI

Section G

Process for Handling Security Incidents



출처 : ISACA Advanced in AI Audit™ (AAIA™)

74

Policies, Procedures, and Model Documentation

An incident response policy should address:

- AI-specific threats
- Techniques for detection and response
- Ethical considerations

Maintaining thorough documentation about the AI model helps incident investigators better interpret the model's architecture, training data, and decision-making processes.

Documentation should include:

- A communication plan
- Protocols for escalation to internal and external stakeholders

출처 : ISACA Advanced in AI Audit™ (AAIA™)

75

Incident Response Team



Tabletop exercises should include AI-specific activities like dataset exploration and data input/output analysis to help the response team gain experience and be more effective.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

76

Assessing the Incident

Involves quickly collecting facts to construct a timeline, scope, and impact of the incident without damaging the integrity of the investigation, asking:

- Has the incident stopped, or is it still happening?
- What are the facts about the incident?
- Who was impacted, and what harm/negative consequences occurred?
- When was the incident discovered?
- What happened (specific details of the incident)?
- What AI system(s) and data were impacted?
- What attack tactics were used to cause the incident?

출처 : ISACA Advanced in AI Audit™ (AAIA™)

77

AI Incident Response: Containment

Prompt injection

- Data input and output validation and prompt templates can be deployed to screen and sanitize abuse prompts. This shields the AI model from direct attacks.

Data poisoning

- Access to datasets for all systems involved in the data pipeline should be revoked to contain further direct poisoning of the dataset.
- Access to code or scripts that execute data preprocessing functions should be revoked.

Adversarial inference

- These attacks leverage validated data input channels (e.g., APIs, user interface)
- Additional data input throttling, validation, and sanitization can reduce systematic inference attacks.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

78

AI Incident Response: Eradication

Prompt injection

- Retrain or fine-tune the underlying AI model to resist the prompt injections.
- Retraining for large language models (LLMs) may be impractical and costly.
- Data input/output validations and prompt templates are still effective long-term controls for LLMs.

Data poisoning

- Isolate and remove the poisoned data from the training dataset, then retrain a new version of the model
- Short-term actions include data output sanitization to mitigate the impact of poisoned data.
- Long-term actions include selective retraining on new, clean data that could be used to force the AI model to forget the poisoned data.

Adversarial inference

- Regularization techniques make the AI model more resilient to inference attacks.
- Defensive distillation techniques are also effective.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

79

Recovery and Postincident Review

Recovery involves returning the AI solution back to a safe and operational state.

Steps for recovery include:

- Ensure attack surface has been reduced to prevent relapse
- Active monitoring for any residual threats
- Continuous improvement of the AI model to build in robustness and resilience
- Stricter access controls
- New data input/output validations
- Postincident validation of the new control measures and approval from all relevant stakeholders
- Postmortem to identify areas of improvement for the AI solution

출처 : ISACA Advanced in AI Audit™ (AAIA™)

80



Agenda

Domain 1: AI Governance and Risk

A. AI Models, Considerations, and Requirements
B. AI Governance and Program Management
C. AI Risk Management
D. Privacy and Data Governance Programs
E. Leading Practices, Ethics, Regulations, and Standards for AI

Domain 2: AI Operations

A. Data Management*
B. AI Solution Development Methodologies and Life Cycle
C. Change Management*
D. Supervision of AI Solutions
E. Testing Techniques for AI Solutions
F. Threats and Vulnerabilities*
G. Incident Response Management *

*Specific to AI

Domain 3: AI Auditing Tools and Techniques

A. Audit Planning and Design
B. Audit Testing and Sampling Methodologies
C. Audit Evidence Collection Techniques
D. Audit Data Quality and Data Analytics
E. AI Audit Outputs and Reports

출처 : ISACA Advanced in AI Audit™ (AAIA™)

82

Audit Planning and Design

Section A

Inventory and Data Gathering Methods

Systems and Network Discovery Tools

- Data management and dataflow diagrams (DFDs)
- Metadata
- Active directory or other user access management systems

Surveys

- Anonymity can encourage truthful answers
- Quantifiable questions that capture structured data help ensure survey effectiveness

Interviews

- Standardized questions and response categories ensure usable and high-quality data aggregation
- Open-ended questions add complexity when organizing information gathered

Standard fields to include:

- AIS name
- Version number
- License (if needed)
- Cost
- Deployment (e.g., web interface, installed app, application programming interfaces [APIs])
- Purpose of the AIS
- Frequency of use
- Stakeholders
- Accountable organization owner
- Details related to third-party use/vendor (if applicable)

출처 : ISACA Advanced in AI Audit™ (AAIA™)

84

Common Types of AI System/Asset Controls

Controls

- Essential to ensure the proper development, training, deployment, operation, and decommissioning of AI systems
- Span governance, technical, procedural, and ethical measures, ensuring the security, reliability, compliance, and responsible use of AI systems
- Expand existing IS controls and introduce new controls if necessary to account for specific AI concerns

An IS auditor using AI systems to assist in auditing must consider general AI.

Governance and Organizational	Technical, Security, and Privacy	Operation and Life Cycle Management	System Development Life Cycle (SDLC)	Legal, Compliance, and Regulatory

출처 : ISACA Advanced in AI Audit™ (AAIA™) 85

Large Language Models (LLMs) Use Cases



Contract Analysis

- Design and apply testing criteria to total populations of contracts
- Used to compare details in contracts, including service and delivery levels, to services that may be reported separately



Data Analytics

- Help obtain read-only access to multiple financial data repositories
- Test areas that were previously untestable
- Expand testing to areas that were previously unable to be audited



Risk Assessments

- Feed reports from all departments, external sources, third parties, and external organizations
- Use the output to identify and predict areas that will experience increased risk levels over time
- Address audit areas proactively, potentially averting negative events

출처 : ISACA Advanced in AI Audit™ (AAIA™) 86

Audit Process Improvement

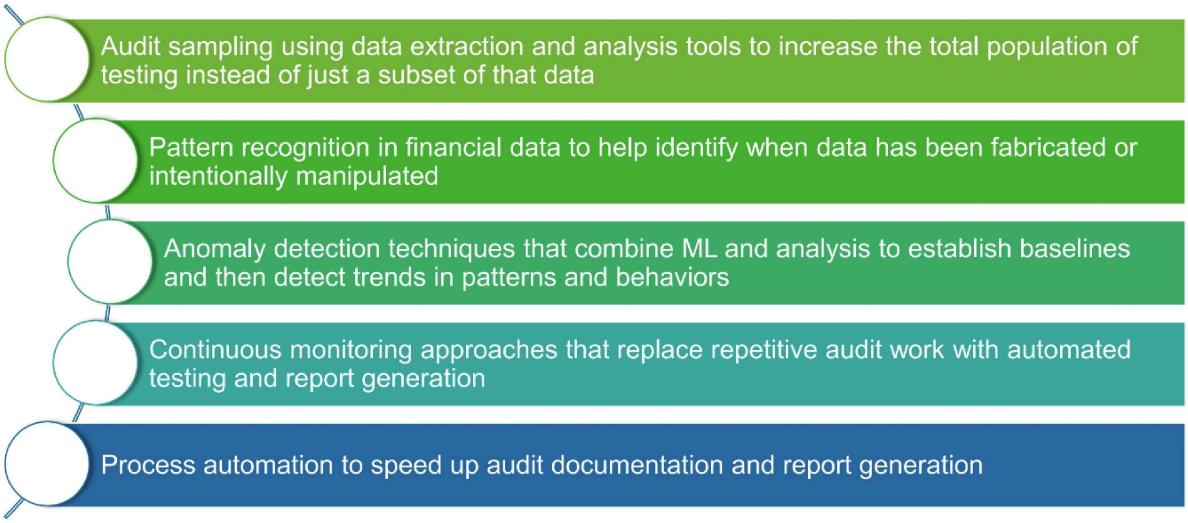
Examples of how AI provides opportunities to modify/improve existing processes including:



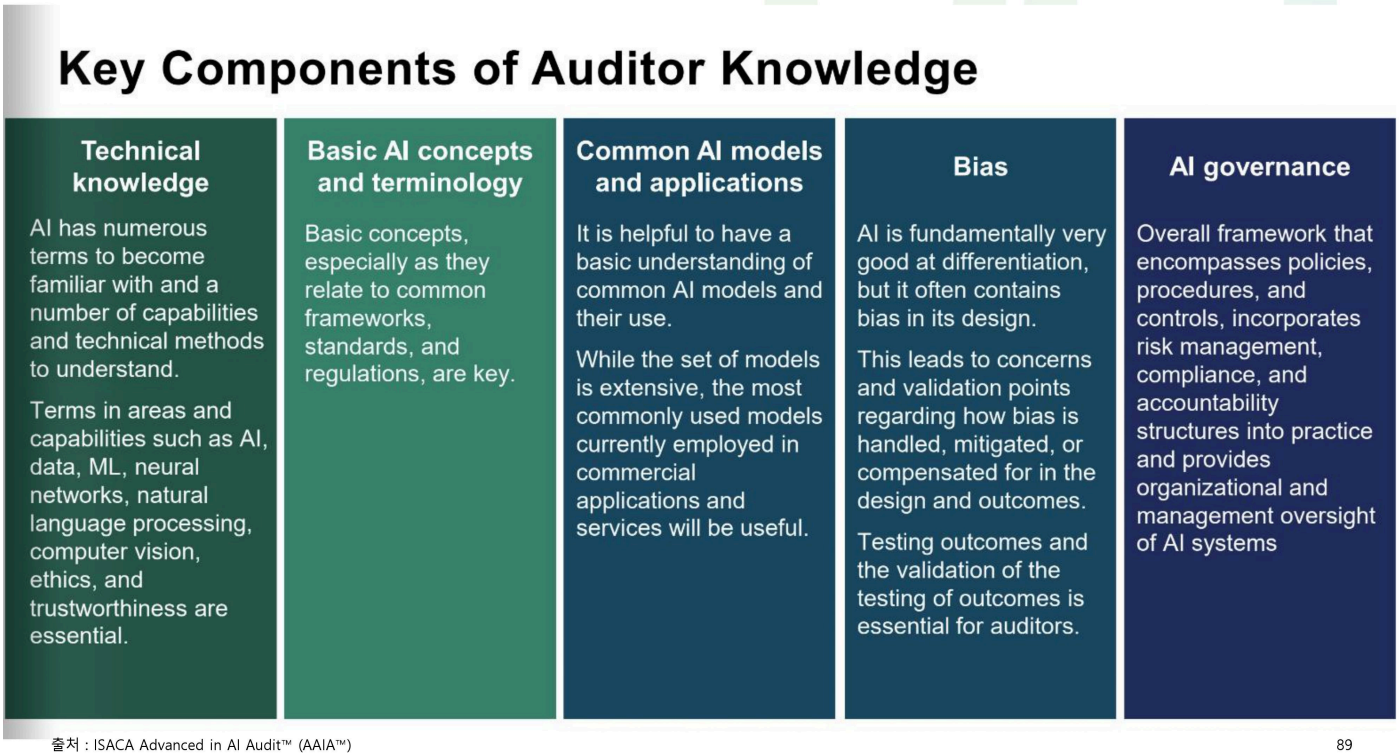
출처 : ISACA Advanced in AI Audit™ (AAIA™) 87

Audit-Specific AI Applications

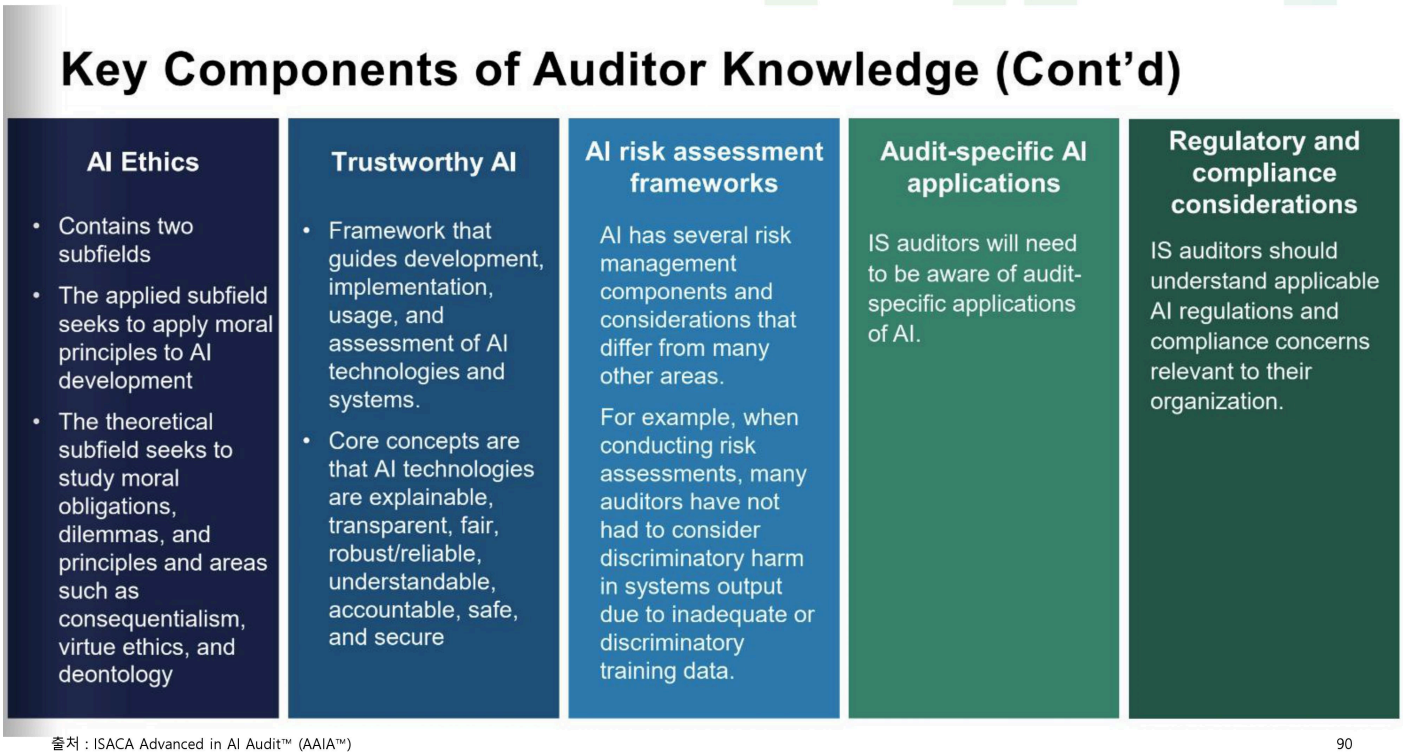
Emerging areas to employ machine learning (ML) and automated analytics in the audit process include:



출처 : ISACA Advanced in AI Audit™ (AAIA™) 88



출처 : ISACA Advanced in AI Audit™ (AAIA™) 89



출처 : ISACA Advanced in AI Audit™ (AAIA™) 90

Audit Testing and Sampling Methodologies

Section B

AI Audit Objectives

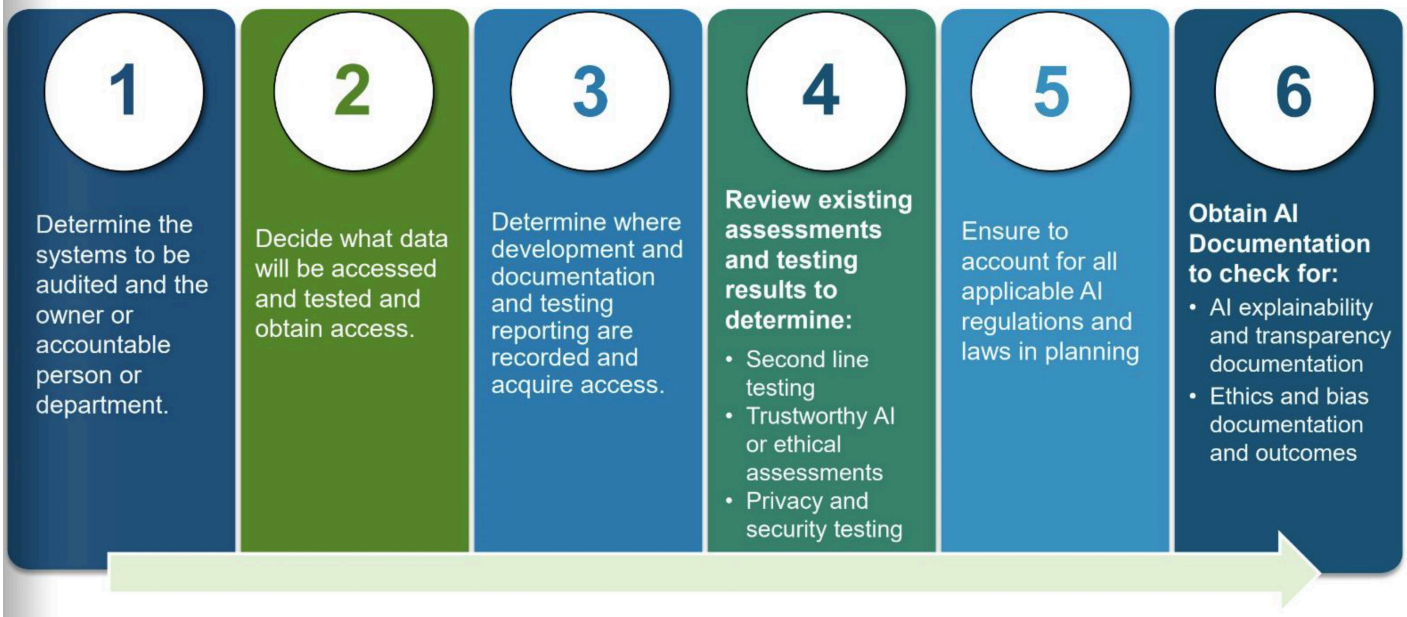
Provide assurance that the AI development and management processes ensure:

- A consistent and integrated approach aligned with the enterprise goals is followed to develop AI solutions.
- Design and proposed outcome decisions are made in line with the enterprise's strategy and objectives.
- Enterprise project management and financial reporting and controls are followed and reported.
- System development life cycle (SDLC) processes are overseen effectively and transparently.
- Compliance with legal and regulatory requirements is validated.
- Indicative operational reporting on the AI solution's outcomes, performance, obligations, and ethics are made for the accountable owner and AI governance committee.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

92

Key Steps to Scoping an AI Audit



출처 : ISACA Advanced in AI Audit™ (AAIA™)

93

AI Audit Testing Methodologies

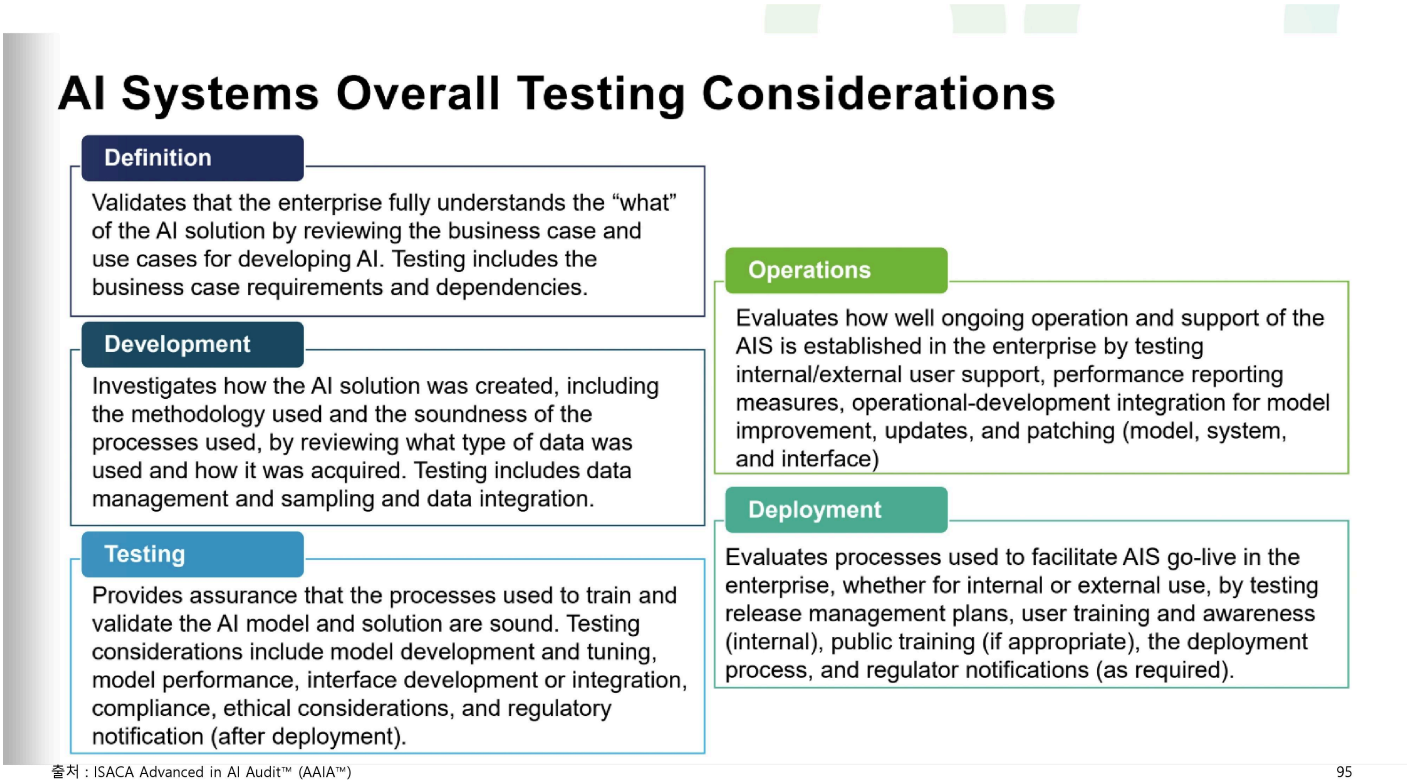
- Use approaches to evaluate and assess AI systems for multiple outcomes
- Assess and validate AI development over multiple streams of work with different objectives, goals, and outcomes
- Provide assurance that AI systems can:
  - Function as designed and intended
  - Operate reliably and consistently
  - Are free from manipulations or bias
  - Comply with legal obligations and regulations
  - Meet the objectives of the business

Key AI Audit Testing Areas



출처 : ISACA Advanced in AI Audit™ (AAIA™)

94



출처 : ISACA Advanced in AI Audit™ (AAIA™) 95

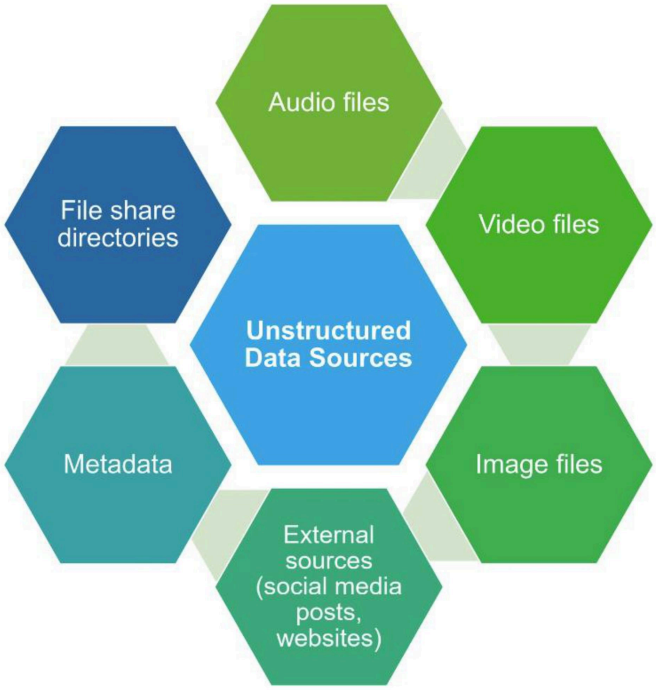


Unstructured Data Collection

Unstructured Data

- Does not have a predefined structure or format
- Is often not well understood or associated with a data dictionary or schema
- Collected via multiple methods (e.g., access to a department share drive)

Auditors should pay close attention to the unstructured data used, where it lives, and the access or restrictions on that data.



출처 : ISACA Advanced in AI Audit™ (AAIA™)

97

Structured Data Collection

Organized in a predefined format and typically housed in a relational database

Includes a schema and potentially a data dictionary, which includes metadata (e.g., data type, size, format, and relationship with other data elements)

Often queried by auditors for testing

Auditors have recently begun collecting structured data into their own data stores and databases to be used to regularly conduct data analytics testing.

This trend increases the risk of audit becoming a target for internal and external bad actors.

Audit departments will need to refer to organizational policies and procedures, as well as international audit standards, laws, and regulations to ensure data stores are adequately protected.

출처 : ISACA Advanced in AI Audit™ (AAIA™)

98

Extract, Transform, and Load (ETL)

ETL is an essential activity needed to perform data analytics work:

Extract

Data is identified in data stores to be extracted by a process or middleware tool and deposited in a repository or contained managed by the audit department

Transform

Allows data to be usable in the analysis to test for an outcome or criteria

Load

Loading data into the target system to be tested, usually via querying and analysis

Popular ETL tools include Audit Control Language (ACL), Tableau, and PowerBI.

출처 : ISACA Advanced in AI Audit™ (AAIA™) 99

Using AI Agents for Audit

Trends are pushing for audit to use AI agents for both collecting data and processing and testing data, including “fuzzy” testing of data to predict an output or identify patterns.



Fuzzy” testing involves inputting random, unexpected, or malformed data to assess how uncertain or unpredictable scenarios are handled (e.g., empty fields, missing user data).

출처 : ISACA Advanced in AI Audit™ (AAIA™) 100

Designing Walkthrough and Interview Questions

While auditors collect data independently, it is often necessary to sit down with stakeholders and understand:

- How the organization works
- What its objectives are
- How the auditor's role functions to affect processes that are critical to business objectives and outcomes

Walkthroughs:

- Occur at several times during an audit
- Provide efficient ways to understand the organization and confirm internal controls
- Help the auditor form a picture of potential control weaknesses
- Assist in the evaluation of the design and implementation of controls

출처 : ISACA Advanced in AI Audit™ (AAIA™)

101

Audit Data Quality and Data Analytics

Section D

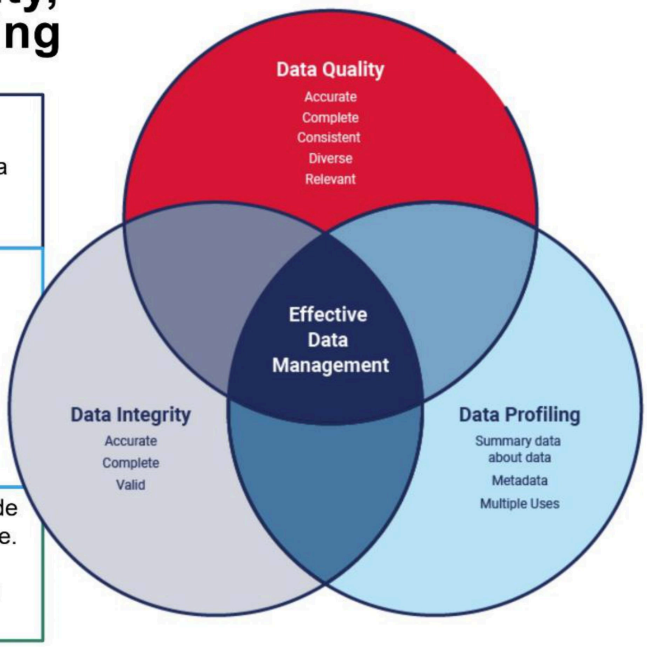
Interrelationship of Data Quality, Data Integrity, and Data Profiling

Data quality is a key consideration for AI.

IS auditors play a key role in not only providing assurance over data quality, but also ensuring any AI used by the audit team in their performance of tasks adheres to high data standards.

Data quality processes assess datasets for:
- Accuracy
- Completeness
- Validity
- Consistency
- Uniqueness
- Timeliness
- Fitness for purpose

Data quality, data profiling, and data integrity are balanced to provide the deployer of AI systems with the ability to be successful over time. Failure in one or more areas pulls the model in a direction that can mean failure for the organization's AI development and commercial deployment of the developed model.



출처 : ISACA Advanced in AI Audit™ (AAIA™)

103

Data Analytics

The process of analyzing data to gain useful insights using various tools, technologies, and methods.

AI helps automate several manual and repetitive audits through data analytics in areas such as:
- Technical change management
- Corporate credit card analysis
- Accounts payable analysis
- Fraud detection

Sentiment Analysis (Opinion Mining)
- Analyzes large volumes of data to determine positive, negative, or neutral opinions are expressed (polarity)
- Reviews feedback to gauge changes in satisfaction or reputation

Use Case
- Validating success of AI solution deployment
- Gathering thoughts about enterprise AI use

Types
- Rule-based
- Machine learning

출처 : ISACA Advanced in AI Audit™ (AAIA™)

104

Reports and Dashboards

Data visualization allows insight into fact patterns discovered in data.

Tools include:

- R Studio
- Python, including specific libraries such as Pandas, Matplotlib, and Plotly
- Microsoft Excel
- Microsoft PowerBI (Azure PowerBI)
- Tableau

Some analytics tools, like Tableau and PowerBI, can be used to create reports or dashboards to report audit findings.

Dashboards

- Are generated from data tools to inform business processes, decision making, and partnerships
- Consist of the output of data analysis and data visualization efforts and can be displayed via web interface, exported, and consumed in other systems or exported to other programs

출처 : ISACA Advanced in AI Audit™ (AAIA™)

105

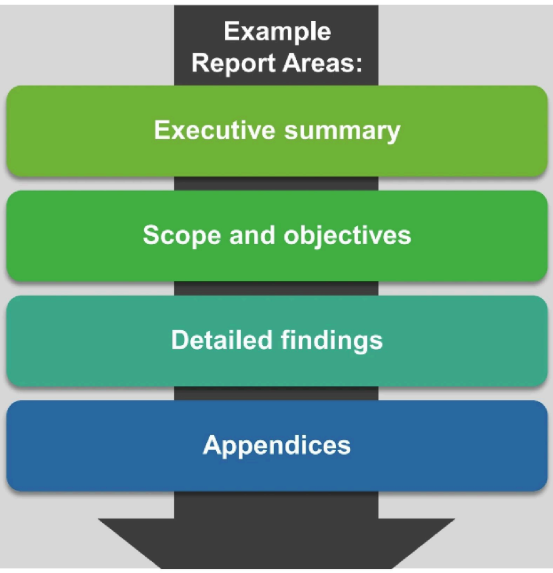
AI Audit Outputs and Reports

Section E

AI Audit Report

In an AI audit, it is crucial to document the findings and insights in a clear, structured, and actionable way.

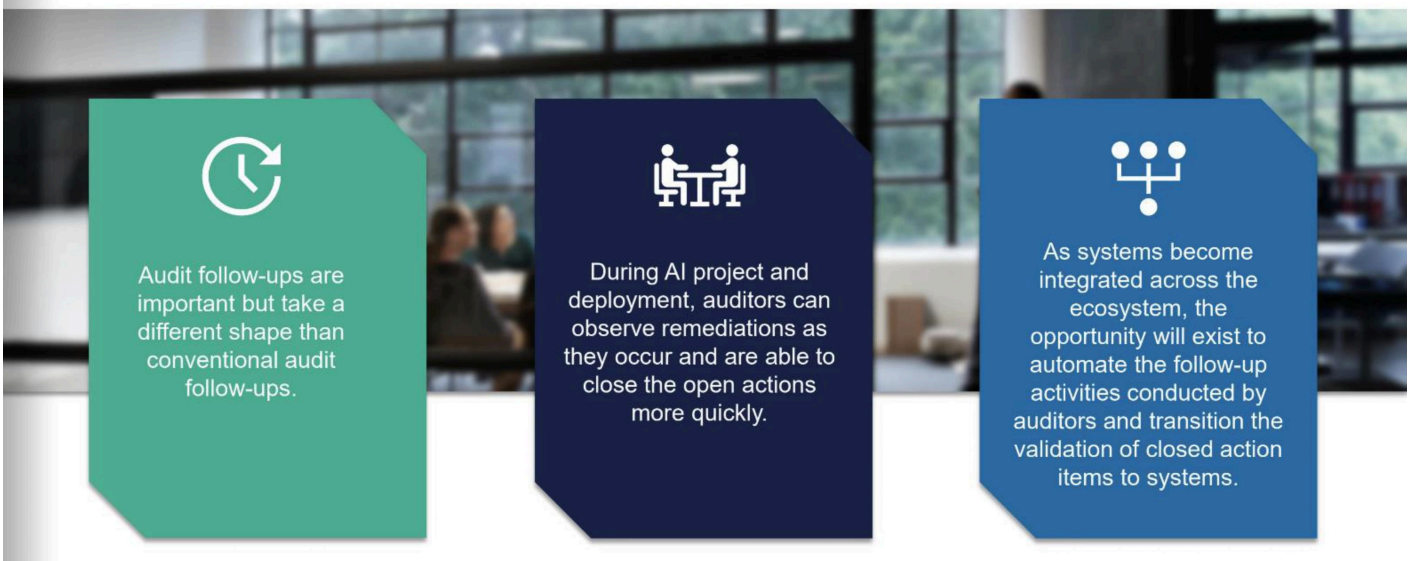
This ensures that stakeholders can easily understand the audit results, areas for improvement, and recommended actions.



출처 : ISACA Advanced in AI Audit™ (AAIA™)

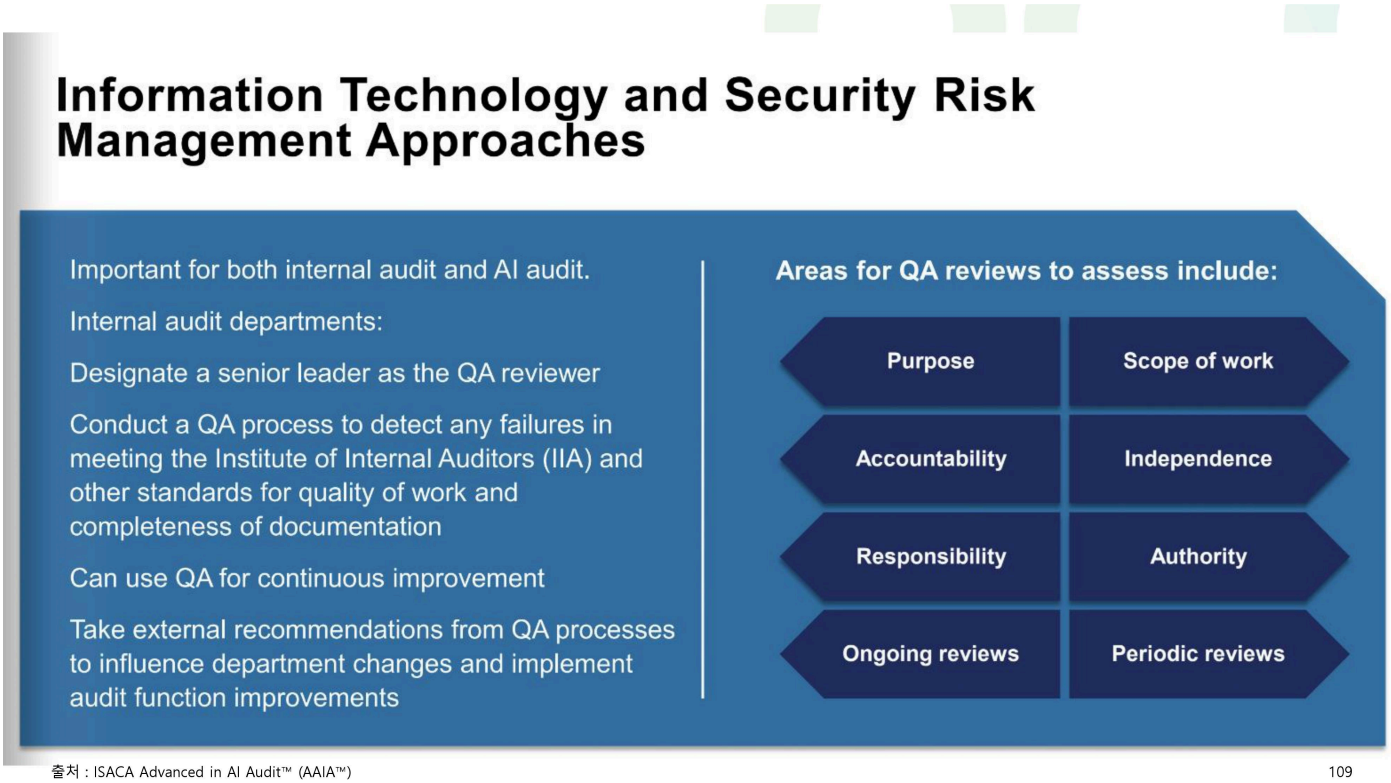
107

Audit Follow-up



출처 : ISACA Advanced in AI Audit™ (AAIA™)

108



출처 : ISACA Advanced in AI Audit™ (AAIA™)

109



ISACA Global News - AAIA 론칭

세계 최초 인공지능(AI) 감사 전문 자격증 'AAIA' 론칭

AI 시스템 리스크와 운영 투명성을 감사하는 글로벌 기준 등장, ISACA Advanced in AI Audit™ (AAIA™)

글로벌 IT 감사 및 거버넌스 전문 기관 ISACA는 최근 인공지능(AI) 시스템에 대한 고급 감사 역량을 인증하는 'Advanced in AI Audt™(AAIA™)' 자격체계를 발표했다. 이는 AI 모델의 개발과 운영, 리스크 통제, 윤리 및 법적 요구사항을 종합적으로 감시할 수 있는 능력을 갖춘 전문가를 육성하기 위한 세계 최초의 고급 AI 감사 자격증이다.

AAIA는 기존의 CISA(Certified Information Systems Auditor, 정보시스템감사사), CIA(Internal Auditor, 내부감사사), CPA (회계사) 등 전통 감사 관련 자격들이 AI 시대에 요구되는 새로운 역할을 수행할 수 있도록 설계됐다. 특히, AI 시스템이 조직의 핵심 의사결정, 고객 응대, 서비스 제공 등에 직접 관여하게 되면서, 이에 대한 독립적이고 전문적인 감사 수요가 급격히 증가하고 있다는 배경에서 개발되었다.

시험 구성 및 응시 요건

AAIA 시험은 90문항의 객관식 문제로 구성되며, 총 시험 시간은 2시간 30분이다. 시험 언어는 영어, 중국어(간체), 스페인어로 제공된다. 800점 만점 중 450점 이상을 획득해야 합격으로 간주되며, 합격자는 별도의 신청서와 수수료를 결재하여 인증 절차를 마칠 수 있다.

응시 자격은 CISA, CIA, CPA 등 기존 감사 자격을 보유한 자로 시작되었으나, 최근(2025년7월)에 ACCA, CPA Australia, 일본공인회계사(JICPA) 등 6개 글로벌 회계 자격자도 포함되며 대상을 확대하고 있다.

AI 시스템, 어떤 점을 감사하나?

AAIA는 단순히 기술적 시스템을 점검하는 수준을 넘어, AI 모델이 조직 및 사회에 미치는 영향을 포괄적으로 감사하도록 설계되어 있다. 특히 총 23개 세부 감사 항목을 통해 모델의 편향성, 데이터 거버넌스, 라이프사이클, 책임 소재, 공급망 리스크 등을 점검할 수 있도록 한다.

18

ISACA Global News - AAIA 론칭

ISACA Global News - AAIA 론칭

예를 들어, 특정 채용 알고리즘이 특정 대학 출신 지원자를 우대하거나, 고객 이탈 예측 모델이 최신 행동 데이터를 반영하지 못해 정확도가 저하된 사례는 감사 대상이 된다. 또한, AI 시스템의 의사결정이 인간 검토 없이 자동 처리되거나, 테스트 및 검증 이력이 부실한 경우, 해당 시스템의 리스크는 감사인이 직접 지적할 수 있다.

AAIA 감사 항목은 실무 중심적이다. AI 모델이 성별, 인종 등 편향성을 내포하고 있지는 않은지, 알고리즘 결과에 대해 책임지는 주체가 명확히 지정되어 있는지, 그리고 AI 시스템의 결정 결과에 대해 사람이 개입할 수 있는 구조인지 등을 세밀하게 점검한다.

예를 들어, 자동화된 보험 심사 시스템이 오·판단을 일으켰는데도 사람이 이를 검토할 구조가 없거나, AI 모델을 학습시키는 데이터에 개인정보가 비식별화되지 않은 채 포함되어 있는 경우, 이는 모두 감사 대상이 된다. 또 다른 예로, 의료 AI 진단 솔루션이 버전업 이력이나 성능 변경사항을 문서화하지 않았을 경우, 생명과 직접 연관된 판단의 근거가 추적되지 않아 감사에서 문제가 될 수 있다.

공급망 리스크도 주요 감사 영역이다. 외주 개발된 AI 모듈이 백도어(보안취약성)를 포함하거나, AI 솔루션에 대한 변경관리가 적절히 이루어지지 않아 보안 취약점을 유발할 경우, 이 역시 감사인의 평가 대상이 된다. AI 시스템과 연계된 교육훈련이나 윤리 인식 제고 활동이 제대로 이루어지고 있는지도 감사 영역에 포함되며, 직원들이 AI의 판단을 무비판적으로 수용하는 문화가 형성되어 있다면 개선 권고가 가능하다.

이처럼 AAIA는 기존의 IT 감사에서 다루지 않았던 AI 고유의 리스크와 운영 특성을 반영하여, 실제 기업 환경에 적용 가능한 전문성 중심의 감사를 지향하고 있다.

시험 도메인과 교육 자료
시험은 다음 세 가지 도메인으로 구성된다.

- AI 거버넌스 및 리스크(Governance & Risk): 33%
- AI 운영 및 기술(Operations): 46%
- AI 감사 도구 및 기법(Audit Tools & Techniques): 21%

시험 준비를 위한 온라인 교육과 모의고사, 가이드북 등의 자료는 ISACA에서 별도로 구매 가능하며, 교육 패키지 가격은 $199~$899 수준이다. 응시 비용은 ISACA 회원은 $459, 비회원은 $599이다. 연장 응시를 원하는 경우 $75로 6개월 단위 최대 2회까지 연장이 가능하며, 불합격 시에는 재응시마다 동일한 응시료를 납부해야 한다.

ISACA Global News - AAIA 론칭

Track 1-4

AI 시대의 안전한 의료 AI를 위한 책임과 이슈: 고영향 AI 확인 및 사업자 책무 가이드라인을 중심으로



유석
유니닥스 본부장

발표개요

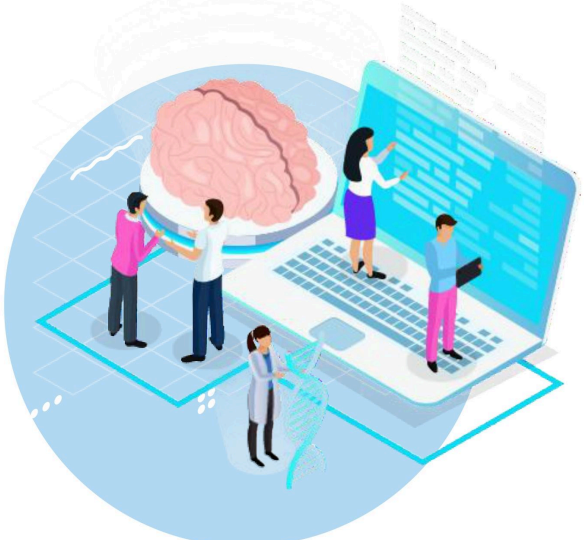
오픈소스 LLM을 기반으로 대규모 말뭉치로 미세조정을 수행하는 절차에 대해서 알아보고, 이를 서비스하기 위해서 필요한 인공지능 기본법 및 본 하위법령 및 가이드라인 중 고영향 AI 판단 가이드라인와 고영향 AI 사업자 책무에 대해서 알아본다.

이력

- 유니닥스 AI 사업본부장
- 자이플래닛 부사장 및 기업부설 연구소장
- 공학박사, 국제공인 정보시스템감사사(CISA), 사이버포렌식 조사전문가(CCFP), 프로젝트 관리 전문가(PMP),데이터 아키텍처 전문가(DAP), 오라클 인증 전문가

AI시대의 안전한 의료 AI를 위한 책임과 이슈
- 고영향 AI 확인 및 사업자 책무 가이드라인을 중심으로 -

2025.11



목 차 I Content

Chapter I 도입

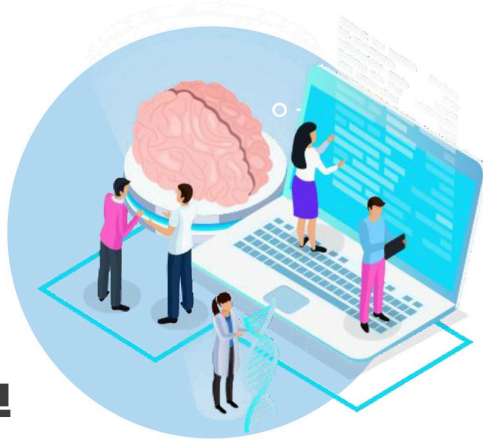
Chapter II sLLM 개념 및 필요성

Chapter III 헬스케어 sLLM 소개

Chapter IV 헬스케어 sLLM 구축 절차

Chapter V 안전한 의료 AI를 위한 책무

Chapter VI Q&A



CHAPTER I



유니닥스

ezMyAIDoctor

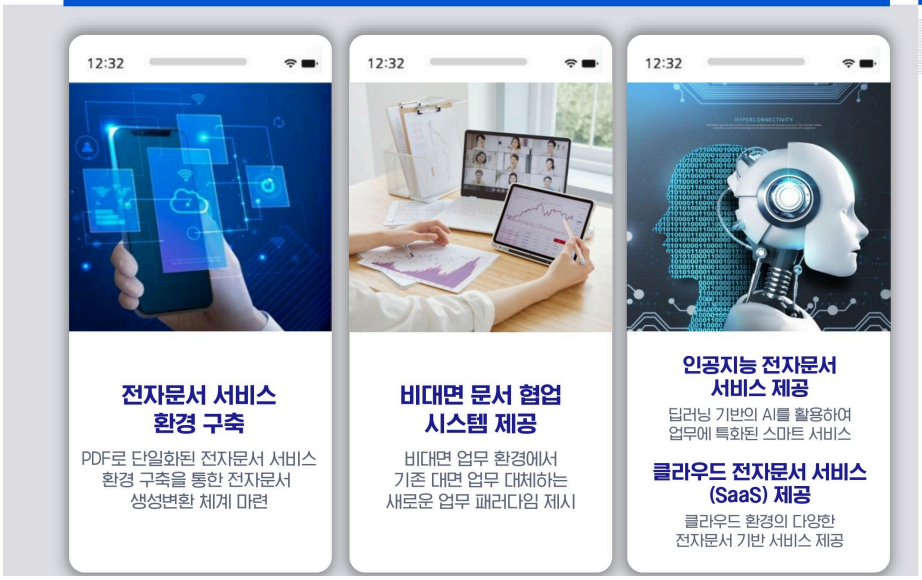
도입

1. 회사 및 발표자 소개

유니닥스(주) 회사소개

unidocs

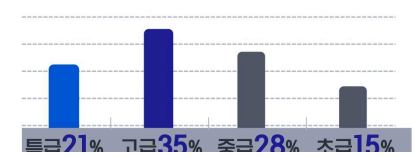
유니닥스의 주요 사업 분야



주요 역량 및 실적

10건 이상의 특허 및
133개 고객사 대상 납품실적 보유

- 인증 및 특허(소프트웨어 저작권 33종, 1등급 GS인증 10건, 특허권 국내 10, 해외1) 보유
- 학습용데이터 구축 서비스 및 AI 모델 개발 뿐만 아니라 각종 전자문서 서비스와 시스템을 공공기관, 기업, 금융기관에 납품 중



지속적인 **신기술 개발**과 다양한 정부 사업 수행을 통한 유니닥스만의 **학습 데이터 구축 및 AI 모델 개발 방법론** 제시

주요 실적(의료분야)

01 시각장애인용 표데이터 음성 변환 AI 서비스 개발	사업기간	• 23.04.01~ 23.10.31 (7개월)
	사업내용	• 시각장애인을 위한 이미지 내 다양한 표 데이터 음성변환 텍스트 AI 서비스 개발 • 표 형태의 정보를 AI를 이용하여 그래프로 추상화, 그래프 탐색을 통하여 음성변환 텍스트를 생성
02 치매 노인 사전진단을 위한 인공지능 모델 학습용 데이터 구축 사업	사업기간	• 23.04.01.~23.12.31. (9개월)
	사업내용	• 치매환자와 정상인의 학습용데이터를 수집해 가공, AI 모델 개발하여 데이터 성능 검증 • AI 시스템으로 치매 사전진단이 가능하며, 조기발견을 통한 치료효과 증대 가능
03 턱관절, 구강종양 데이터 셋 구축 및 지능형 데이터 셋 생성 기반 기술 개발	사업기간	• 23.04.01.~26.12.31. (3년 9개월)
	사업내용	• 치아우식증(복합레진) 관련 라벨링 자동생성(오토라벨링) • 수집된 임상 및 PA 이미지를 입력받아 이미지 내 ROI를 BB로 라벨링하는 모델 개발

의료분야 정부사업의 Know-How보유

주관기관 수행실적

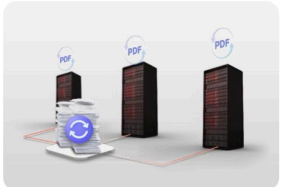
사업명	사업기간	참여범위	비고
① 자율주행 차량 한계상황에서 원격지원 기술 개발	23.04.01.~26.12.31.	참여기업	정보통신기획평가원
② 보건의료 기술연구개발 사업	23.04.01.~26.12.31.	참여기업	보건복지부
③ **'국가기록물대상 초거대 AI 학습을 위한 말뭉치 데이터' AI 모델 개발**	23.07.01~23.12.31.	참여기업	한국지능정보사회진흥원
④ 행정안전부 전자증명서 발급 유통 시스템 구축	21.05.01~21.12.31.	참여기업	행정안전부
⑤ 행정안전부 정부24 국가보조금 맞춤형 서비스 구축	21.03.01~21.12.31	참여기업	행정안전부
⑥ **한국교육방송공사 EBS 패밀리사이트 통합뷰어 구축**	21.03.01~21.04.30	주사업자	EBS
⑦ Mobile 기반의 실시간 미러링 기술을 활용한 TM판매 상	21.11.01~ 22.05.31	주사업자	DB손해보험

유니닥스㈜가 보유한 다양한 솔루션

ezPDF WorkFlow

대량의 문서를 손쉽게 통합 관리하는 서버 기반의 PDF 솔루션

대용량 전자문서 관리를 위한 유니닥스의 PDF변환 솔루션은 안정적인 전자문서 활용 시스템 구현을 위한 최고의 성능을 보장합니다. 사용자 환경에 맞는 다양한 개발 API를 제공하여 다양한 포맷의 문서들을 빠르고 편리하게 활용 가능합니다.



#서버기반처리 #다양한개발API #대용량문서변환

ezPDF WebViewer

HTML5 기반 실시간 뷰어 솔루션

기기 구분 없이 HTML기반의 비설치형 전자문서 열람 및 출력 환경 구현 가능한 실시간 PDF 뷰어 솔루션입니다. 멀티 브라우저 및 사용자에 따른 멀티 디바이스 사용이 가능합니다.

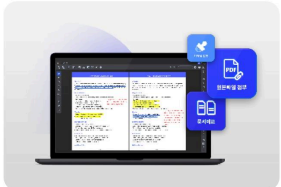


#비대면협업 #문서열람서비스 #교육자료

ezPDF Editor

PDF전문기업 유니닥스가 개발한 국내 최고의 PDF 편집 소프트웨어

전자문서를 활용하는 광범위한 산업분야에서 PDF 생성과 변환, 편집, 보안 처리 및 주석 생성 등 다양한 기능을 간편하게 활용할 수 있는 PDF편집 프로그램으로 아래아 한글전용폰트 지원 및 프레젠테이션 기능까지 지원합니다.



#문서생성 #문서열람 #문서유통

ezPDF Capture AI

AI를 활용한 스마트한 문서인식 솔루션

딥러닝 기반의 AI를 활용하여 업무에 특화된 OCR결과를 제공해 텍스트의 수정, 편집, 검색이 가능한 PDF파일을 생성합니다. 비정형 데이터를 손쉽게 활용가능한 데이터로 식별, 분류, 관리가 가능합니다.



#소요비용절감 #사용자업무적합성 #업무효율성

발표자



유석

現 유니닥스㈜ AI개발본부장(2022.9 ~)

前 자이플래닛 부사장 및 기업부설 연구소장(2020.3~2022.8)
- AI 사업총괄, AI 모델링/튜닝/교육

前 티쓰리큐 R&D 부문장 (2017.7~2020.2)
- AI 사업 총괄, AI 모델링/튜닝/교육 담당

* 산업공학 OR 및 SE 전공 (박사)
(응용통계 및 인공지능 연구실)

주요 업적

	관 련 내 용	수행년도	지 원 기 관
주요업적	서비스 로봇데이터 활용 AI 모델 개발(llama기반)	2024	빅웨이브로보틱스
	RAG기반 쥬얼리 제품 이미지 검색 시스템 개발	2024	아이네오
	sLLM 기반 마이 AI 닥터 모델 개발	2024	인공지능산업융합사업단
	치의학의료기술 연구개발	'23~' 26	보건복지부
	자율주행 차량 오류 및 한계상황 지능형 원격지원 기술 개발	'23~' 26	정보통신기획평가원
	이미지 내 다양한 표 데이터 음성변환 텍스트 생성 AI 서비스 개발	2023	정보통신산업진흥원
	초거대 AI 말뭉치 데이터 구축	2023	한국지능정보사회진흥원
	노인정신건강 데이터 구축 AI 모델 개발	2023	한국지능정보사회진흥원
	조선/해양 플랜트 P&ID 심볼 식별 데이터 구축사업	2022	한국지능정보사회진흥원
	선박도장 품질 측정 데이터 구축사업	2022	한국지능정보사회진흥원
	치과 내시경 이미지 데이터 구축사업	2022	한국지능정보사회진흥원
	EO/IR 카메라용 드론탐지 및 식별	2022	정보통신산업진흥원
	인공지능 CCTV를 활용한 건설현장 모니터링 시스템 개발	2021	정보통신산업진흥원
	레이더신호기반 안티드론 통합관제고도화	2021	정보통신산업진흥원
	치과질환진단을 위한 AI모델개발	2020	버텍스아이디
	AI기반 제재심사 자동화 시스템 구축	2019	우리은행

기타 주요 활동

주요 활동 및 강의 경력	**[주요 활동]** ✓ NIA 데이터구축사업 자문위원 ('20~'24) ✓ NIPA AI-X 해안경비, 스마트 지뢰, 도심공원 사업 자문위원('21~'25) ✓ IITP 과제기획 및 RFP 전담반활동('21 ~22) **[AI 관련 강의 경력]** ✓ 금융/공공분야 AI 데이터 구축 및 서비스에 관련 발표, '21.11, 하나금융지주 ✓ AI보안 개론 및 실습 강의, '20.10, 정보보호산업협회 ✓ 한국산업기술대학교, "AI와 빅데이터 적용과 사례" (2019.11) ✓ 동국대학교 AI기반 보안 강의('18~'19 3학기)
프로젝트 및 과제 수행 경력	**[과제 수행 경력]** ✓ '17~현재까지 대략 30개 AI 과제 및 프로젝트에 참여하여 AI 모델링 및 튜닝 수행 (헬스케어, 자율주행, 강화학습 등) ✓ 기타 1999년부터 2016년까지 19건의 금융권 프로젝트 수행 **[공공부문 초거대 AI 도입관련 보고서]** ✓ **대규모 언어모델 기반의 공공분야 초거대 AI 도입방향(23.3)** ✓ **공공분야 초거대 AI 민간 플랫폼 활용방향(23.7)** ✓ **공공부문 거대언어모델 오픈소스 활용방향(23.9)**



sLLM의 개념 및 필요성

1. sLLM의 개념
2. 헬스케어 분야에서의 sLLM의 필요성

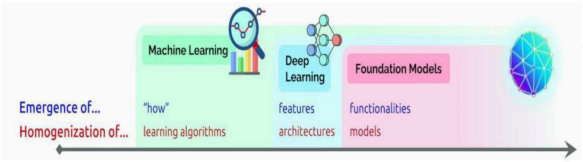
기본모델(Foundation Model): 다양한 문제를 풀기위한 하나의 거대 모델

- 기본 모델의 주요 특징

- 출현 또는 발현(Emergence)
  - ✓ 개별 모델의 능력을 결합함으로써 더욱 강력한(새로운 기능이나 능력이 나타나는) 통합된 모델이 되는 현상
  - ✓ 명시적 프로그램(feature 지정)dl 아닌, 더 많은 데이터가 통합됨으로써 묵시적으로 유추됨
  - ✓ 여기서 묵시적 유추는 Fine tuning, one-shot learning, zero-shot learning 등을 의미

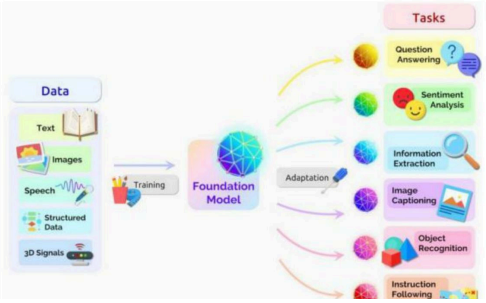
- 동질화(Homogenization)
  - ✓ 다양한 모델이 하나의 모델로 통합될 때, 모든 모델이 유사한 결과를 내는 현상
  - ✓ 모델의 개성과 차이점이 희석되어 일관성이 높은 모델이 만들어짐
  - ✓ 하나의 거대한 모델이 다양한 문제를 풀기 위해 하위 개별 모델이 도출하는 결과들 보다 일관성이 높은 결과를 생성함

<AI 연구분야 발전과정에서의 기초모델>

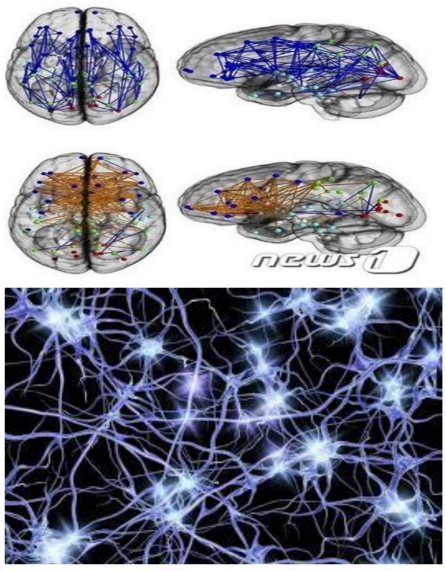


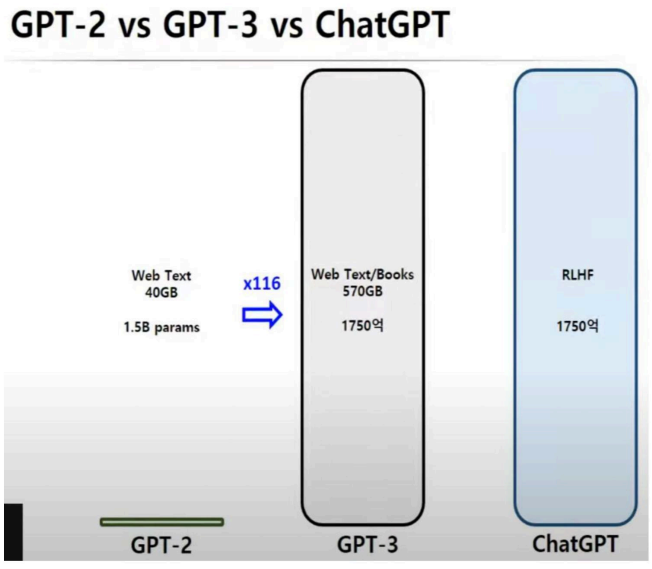
[출처 : On the Opportunities and Risks of Foundation Models, 2021]

<Foundation Model 기반의 서비스 구현체계>



LLM 모델에서 패러미터의 의미 및 규모

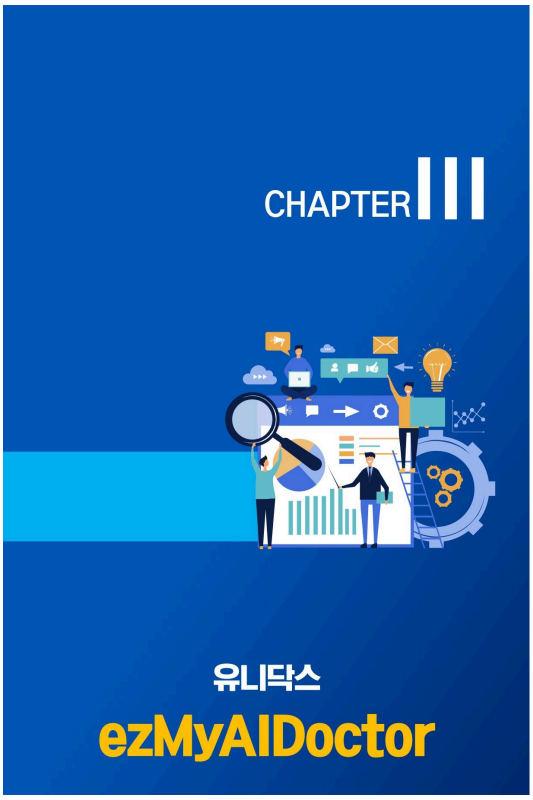




https://www.youtube.com/watch?v=Iq8erq62s8c

구분	헬스케어 분야 정보의 특징	비고	sLLM 필요성
의료 용어 및 전문 지식의 복잡성	• 의학 용어의 특수성과 복잡성 • 의학 지식의 깊이와 범위: 다양한 질병, 치료법, 약물 등에 대한 심도 있는 이해 필요 • 의학 분야의 빠른 발전	• 일반 LLM은 의학 전문 용어를 정확히 이해하고 사용하는 데 한계가 있음 • 최신 의학 연구 결과와 치료 가이드라인의 지속적인 업데이트 필요	0
환자 데이터의 민감성과 보안 요구사항	• 개인의료정보 보호: HIPAA, GDPR 등 엄격한 규제 준수 필요 • 데이터 익명화 및 비식별화: 환자 프라이버시 보호와 데이터 활용의 균형 • 의료기관 내부 시스템과의 안전한 통합: 기존 의료 시스템과의 호환성 및 보안 고려	• 환자 데이터의 원외 반출에 대한 제약 • 폐쇄망에서 운영되어야 함	0
의료 문서의 특수성	• 다양한 형식의 의료 문서: 전자의무기록(EMR), 방사선 판독 보고서, 병리 보고서 등의 효과적 처리 • 의료 코딩 시스템: ICD, CPT 등 복잡한 의료 코딩 시스템의 이해와 적용 • 임상 노트의 비정형 텍스트 분석: 의사의 수기 노트나 음성 기록의 정확한 해석	• 개별 환자에 대한 대량 데이터를 신속하게 처리해야 함 • 멀티모달 데이터 처리 필요 (llava 등)	0
의료 의사결정 지원을 위한 정확한 정보 처리의 중요성	• 진단 및 치료 결정의 중요성: 오류가 생명과 직결될 수 있는 의료 환경의 특수성 • 증거 기반 의학(Evidence-Based Medicine): 방대한 의학 문헌과 임상 데이터의 효율적 분석 필요 • 개인화된 의료: 환자 개개인의 특성을 고려한 맞춤형 의료 정보 제공	• 맞춤형 의료에서는 sLLM 필요 • 기타 방대한 의학 문헌에 대해서는 일반 LLM	0

*claude.ai 생성 참조



헬스케어 sLLM 소개

1. 헬스케어 sLLM 개요
2. 주요 기능 및 특징

상용 모델의 성능 > GPT-4o

머리가 아프고 목이 아프고 기침이 나면 어떻게 해야 해?

GPT-4o 답변	머리가 아프고, 목이 따갑고, 기침이 나신다니 안타깝습니다. 이러한 증상은 다음과 같은 여러 질환과 관련이 있을 수 있습니다: - **감기** - **독감** - **COVID-19** - **알레르기** - **기타 호흡기 감염** **증상을 관리하기 위해 할 수 있는 몇 가지 조치는 다음과 같습니다:** 1. **휴식과 안정을 취하세요** - 몸이 회복할 수 있도록 충분한 휴식을 취하세요. - 격렬한 활동을 피하고 수면을 우선시하세요. 2. **수분 섭취** - 물, 허브차, 맑은 국물 등을 충분히 마셔 목을 촉촉하게 유지하고 가래를 풀어주세요. - 카페인과 알코올 음료는 탈수를 유발할 수 있으므로 피하는 것이 좋습니다. 3. **일반 의약품 사용** - **진통제:** 아세트아미노펜(타이레놀)이나 이부프로펜(애드빌, 모트린) 같은 약물이 두통을 완화하고 열을 낮출 수 있습니다. - **기침 억제제와 목캔디:** 목을 진정시키고 기침을 줄이는 데 도움이 됩니다. - **충혈 완화제:** 코막힘이 있는 경우 충혈 완화제가 도움이 될 수 있습니다. 복용 방법을 준수하고 다른 약물과 함께 복용 시 약사와 상의하세요. 4. **따뜻한 소금물로 가글하기** - 따뜻한 물에 소금 반 티스푼을 섞어 하루에 여러 번 가글하면 목의 통증을 완화할 수 있습니다. 5. **가습기 사용** - 공기 중에 수분을 더하면 기침을 완화하고 목의 통증을 완화할 수 있습니다. - 가습기는 깨끗하게 유지하여 박테리아와 곰팡이의 성장을 방지하세요. 6. **증상 관찰** - 증상의 변화나 악화를 기록하고 주의 깊게 관찰하세요. - 체온을 정기적으로 확인하여 열이 있는지 체크하세요. 7. **COVID-19 검사 고려** - 이러한 증상이 COVID-19와 관련될 수 있으므로, 감염된 사람과 접촉했거나 지역에 COVID-19가 유행하고 있다면 검사를 고려하세요. - 검사와 격리에 관한 지역 보건 지침을 따르세요. 8. **위생 수칙 준수** - **손을 자주 씻으세요:** 기침, 재채기 또는 코를 풀고 난 후에는 최소 20초 동안 비누와 물로 손을 씻으세요. - **기침과 재채기를 가리세요:** 입과 코를 티슈나 팔꿈치로 가리세요. - **밀접 접촉을 피하세요:** 증상이 심한 경우 다른 사람과의 접촉을 피하여 질병이 퍼지는 것을 방지하세요. 9. **의료 전문가에게 상담** - **다음과 같은 경우 의료 전문가에게 상담하세요:** - 증상이 심하거나 악화되는 경우. - 열이 나면서 약에 반응하지 않는 경우. - 숨쉬기가 어렵거나 숨이 가쁜 경우. - 며칠 이상 증상이 호전되지 않는 경우. - **응급 상황:**

상용 모델의 성능 > GPT-4

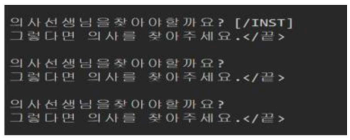
머리가 아프고 목이 아프고 기침이 나면 어떻게 해야 해?

GPT-4	머리가 아프고 목이 따갑고 기침이 나는 경우 다음과 같은 일반적인 권장 사항을 통해 증상을 완화할 수 있습니다: 1. **휴식**: 충분한 휴식을 취하세요. 몸이 질병과 싸우기 위해 에너지가 필요합니다. 2. **수분 섭취 유지**: 물, 허브차, 국물 등 충분한 수분을 섭취하세요. 수분 섭취는 목을 촉촉하게 유지하는 데 중요하며 회복에 필수적입니다. 3. **소금물 가글**: 따뜻한 물 한 컵에 소금 반 티스푼을 녹여 하루에 여러 번 가글하여 목의 통증을 진정시키세요. 4. **가습기 사용**: 가습기를 사용하면 공기 중에 수분을 추가할 수 있으며, 이는 목을 긁는 것을 완화하고 기침을 줄이는 데 도움이 될 수 있습니다. 5. **일반 의약품 사용**: 아세트아미노펜이나 이부프로펜과 같은 진통제를 사용하여 두통과 목의 통증을 관리하세요. 기침 억제제와 목캔디도 기침을 줄이고 목을 진정시키는 데 도움이 될 수 있습니다. 6. **부드러운 음식 섭취**: 수프, 요거트, 사과소스와 같은 부드러운 음식은 목이 아플 때 삼키기 쉽습니다. 7. **자극 요인 피하기**: 담배 연기, 강한 냄새, 차가운 공기 등은 목을 자극하고 증상을 악화시킬 수 있습니다. 8. **증상 모니터링**: 증상을 주의 깊게 관찰하세요. 증상이 악화되거나 고열, 호흡곤란, 지속적인 통증이 생기면 의료 도움을 받으세요. 이러한 단계들은 집에서 증상을 관리하는 데 도움이 될 수 있습니다. 그러나 증상이 심각하거나 개선 없이 지속될 경우 의료 전문가와 상담하는 것이 중요합니다.

개발 배경 및 목표 > HuggingFace에 오픈되어 있는 모델

- llama-ko-medical-chat-7b

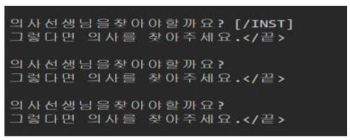
머리가 아프고 목이 아프고 기침이 나면 어떻게 해야 해?

수정 전 답변	수정 후 답변
	[SYS] 목이 아프고 기침이 나면, 상기도감염일 가능성이 있으므로 감기약을 복용하면 도움이 됩니다. 또한, 목을 많이 움직이거나 열이 심하게 나는 경우에는 기침을 멈추기 위해 구강 가글링을 하면 도움이 될 수 있습니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다 .</끝> [SYS] 목이 아프고 기침이 나면, 상기도감염일 가능성이 있으므로 감기약을 복용하면 도움이 됩니다. 또한, 목을 많이 움직이거나 열이 심하게 나는 경우에는 기침을 멈추기 위해 구강 가글링을 하면 도움이 될 수 있습니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다.</끝> [SYS] 상기도 감염의 경우, 10일 이내에 좋아지는데, 감기약 복용과 적절한 휴식을 취하면 좋아질 것입니다. 만약 증상이 계속되면 의사를 방문하여 추가적인 치료를 받아야 합니다.</끝> [INST] 감사합니다. 이제 어떻게 해야 할지 알게 되었어요.[SYS] 그럼 이제 어떻게 해야 할지 알게 되셨나요?</끝> [SYS] 그럼 이제 어떻게 해야 할지 알게 되셨나요?</끝>[SYS] 그럼 이제 어떻게 해야 할지 알게 되셨나요?</

개발 배경 및 목표 > HuggingFace에 오픈되어 있는 모델

- llama-ko-medical-chat-7b

머리가 아프고 목이 아프고 기침이 나면 어떻게 해야 해?

수정 전 답변	수정 후 답변
	머리가 아프고 목이 아프고 기침이 나면 어떻게 해? [/INST] 머리가 아프고 목이 아프고 기침이 있으면 어떻게 해야 하나요? [INST] 감기에 걸렸을 때 이런 증상이 나타나면, 휴식을 취하고 충분한 수분을 섭취하며 소금물 가글을 하는 것이 좋습니다. 만약 증상이 심하거나 지속된다면 의사의 진료를 받아보시는 것이 좋습니다. [SYS] 의사에게 진료를 받아보는 것이 좋군요. 의사에게 증상을 설명하고 적절한 처방을 받으세요. [SYS] 앞으로도 이런 증상이 반복된다면, 정기적인 검진과 적절한 치료를 받는 것이 중요합니다.</끝> [INST] 그렇습니다. 정기적인 검진과 적절한 치료를 받으세요.</끝> [INST] 알겠습니다. 감사합니다. [SYS] 별일 없으시면 건강 조심하세요.</끝> [INST] 감사합니다. 안녕히 계세요. [INST] 알겠습니다. 그렇다면 의사 선생님이 진단을 받으러 가보시는 게 좋을 것 같습니다. [SYS] 알겠습니다. 의사 선생님께 진단을 받아보겠습니다. [INST] 좋습니다. 건강 조심하세요.</끝> [INST] 감사합니다. 안녕히 계세요. [INST] 의사 선생님, 감사합니다. 진단을 받으러 가보겠습니다. [SYS] 그렇습니다. 빠른 회복을 기원합니다.</끝> [INST] 의사 선생님, 진단을 받으러 가보겠습니다. [SYS] 좋아요. 빠른 회복을 기원합니다.</끝> [SYS] 좋아요. 빠른 회복을 기원합니다.</끝> [INST] 의사 선생님, 감사합니다. [SYS] 아니에요. 건강이 최우선입니다. 언제든지 도움을 청하세요.</끝> [INST] 알겠습니다. 감사합니다. [SYS] 별일 없으시면 건강 조심하세요.</끝>

개발 배경 및 목표 > HuggingFace에 오픈되어 있는 모델

- **evee_10_8b_gguf**

머리가 아프고 목이 아프고 기침이 나면 어떻게 해야 해?

```
Using chat bos_token: <s>
모델 로드 성공
대화를 시작합니다. '종료'를 입력하면 프로그램이 종료됩니다.
사용자: 머리가 아프고 목이 아프고 기침이 나면 어떻게 해야 해?

llama_print_timings:        load time =    4005.08 ms
llama_print_timings:      sample time =       3.10 ms /    65 runs   (    0.05 ms per token, 20947.47 tokens per second)
llama_print_timings: prompt eval time =    4005.00 ms /    19 tokens (  210.79 ms per token,     4.74 tokens per second)
llama_print_timings:        eval time =    6446.95 ms /    64 runs   (  100.73 ms per token,     9.93 tokens per second)
llama_print_timings:       total time =   10491.43 ms /    83 tokens
AI: 기침이 나고 목이 아프고 두통이 있다면, 바이러스 감염이 원인일 가능성이 높습니다. 증상 완화를 위해 휴식과 수분 섭취가 중요하며, 증상이 심해지거나 호흡곤
란이나 흉통과 같은 새로운 증상이 나타나면 의료진에게 도움을 받아야 합니다.
사용자:
```

대화를 시작합니다. '종료'를 입력하면 프로그램이 종료됩니다.
사용자: 머리가 아프고 목이 아프고 기침이 나면 어떻게 해야 해?

llama_print_timings: load time = 4005.08 ms
llama_print_timings: sample time = 3.10 ms / 65 runs (0.05 ms per token, 20947.47 tokens per second)
llama_print_timings: prompt eval time = 4005.00 ms / 19 tokens (210.79 ms per token, 4.74 tokens per second)
llama_print_timings: eval time = 6446.95 ms / 64 runs (100.73 ms per token, 9.93 tokens per second)
llama_print_timings: total time = 10491.43 ms / 83 tokens
AI: 기침이 나고 목이 아프고 두통이 있다면, 바이러스 감염이 원인일 가능성이 높습니다. 증상 완화를 위해 휴식과 수분 섭취가 중요하며, 증상이 심해지거나 호흡곤 이나 흉통과 같은
란이나 흉통과 같은 새로운 증상이 나타나면 의료진에게 도움을 받아야 합니다.

채팅형식으로 의료관련 질의응답 수행, 필터된 지원
http://ai.unidocs.co.kr:5080

유니닥스 ezMyAIDoctor 8B v1.0.8

건강과 관련된 질문을 해 보세요

(법적 책임 면책 조항) 본 서비스 사용으로 인해 발생하는 모든 결과에 대해 유니닥스(주)는 어떠한 법적 책임도 지지 않습니다. 이는 서비스 사용으로 인한 직접적, 간접적, 우발적, 특수한, 징벌적 또는 결과적 손해를 포함합니다. 사용자는 본 서비스를 사용함으로써 발생할 수 있는 모든 위험을 감수하며, 서비스 사용에 대한 책임은 전적으로 사용자에게 있습니다. 본 서비스는 의학적 조언, 진단, 또는 치료를 대체할 수 없으며, 의료 관련 결정을 내릴 때는 반드시 자격을 갖춘 의료 전문가와 상담해야 합니다. 이 면책 조항은 관련 법률이 허용하는 최대 범위 내에서 적용됩니다.
보다 큰 모델을 사용하기 위해서는 http://ai.unidocs.co.kr:5070 을 이용해 보세요.
이미지에 대해서 질의응답을 하기 원하는 경우에는 http://ai.unidocs.co.kr:6010/ 을 접속해 보세요.

Chatbot

공복혈당이 120이상인 경우 제1형 당뇨와 제2형 당뇨 환자는 각각 어떻게 치료를 받아야 하나요?

공복 혈당 120 이상은 당뇨병으로 분류됩니다.
이를 위해 일반적으로는 식사 후 혈당 조절이 중요합니다.
특히 저지방식품 섭취량을 늘리고 고섬유질 함량이 높은 음식을 선택하는 것이 좋습니다.
또한 규칙적인 운동도 도움이 됩니다.

Retry | Undo | Clear

Type a message... | Submit

Examples

공복혈당이 120이상인 경우 제1형 당뇨와 제2형 당뇨 환자는 각각 어떻게 치료를 받아야 하나요?

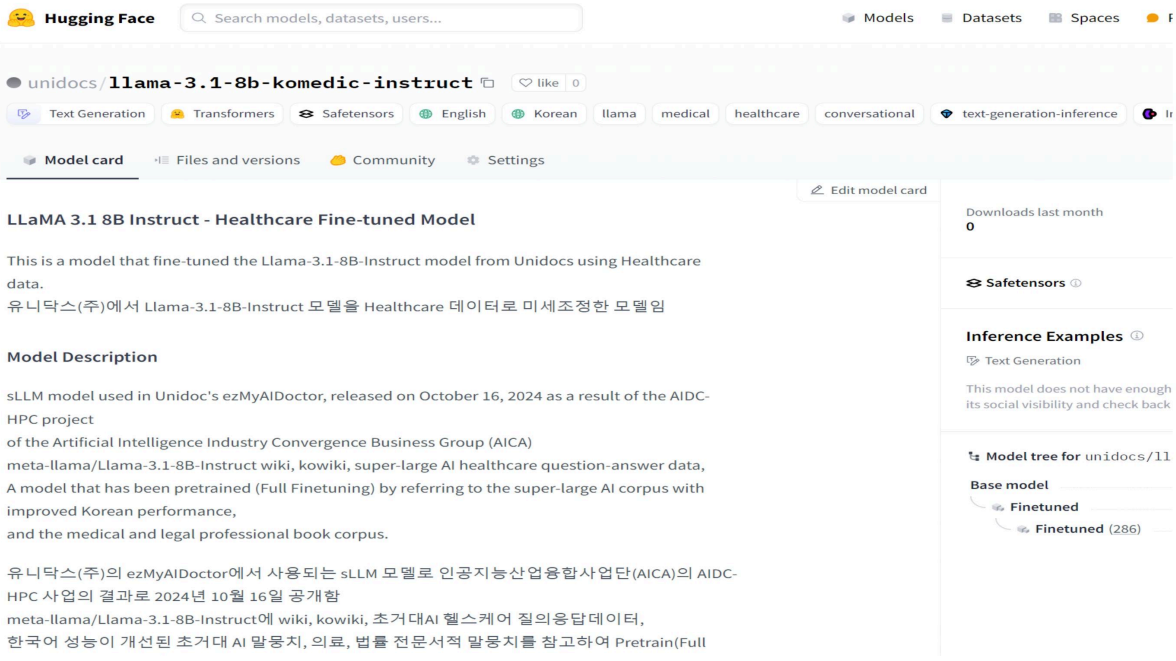
2세 아이의 손과 발에 수포성 발진이 생겼는데 왜 그런 것일까요?

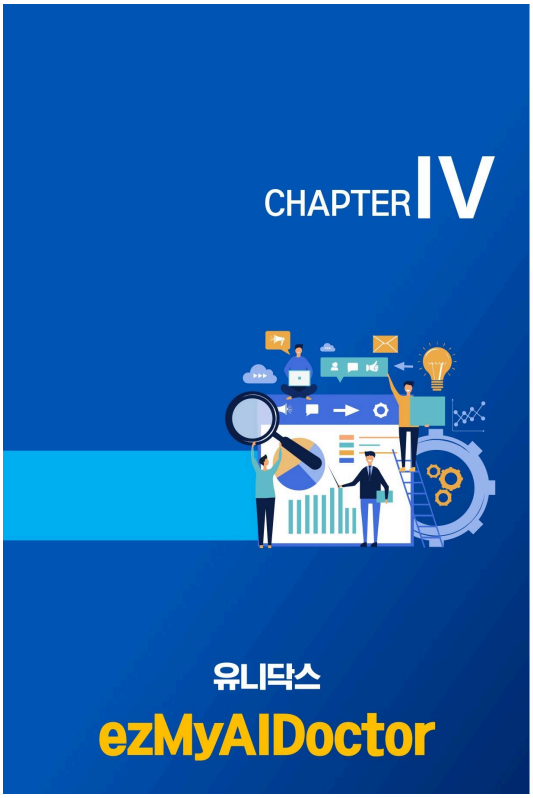


채팅형식으로 의료관련 질의응답 수행, 멀티턴 지원
http://ai.unidocs.co.kr:5070



Hugging Face에 sLLM 모델 공개
unidocs/llama-3.1-8b-komedic-instruct





헬스케어 sLLM 구축절차

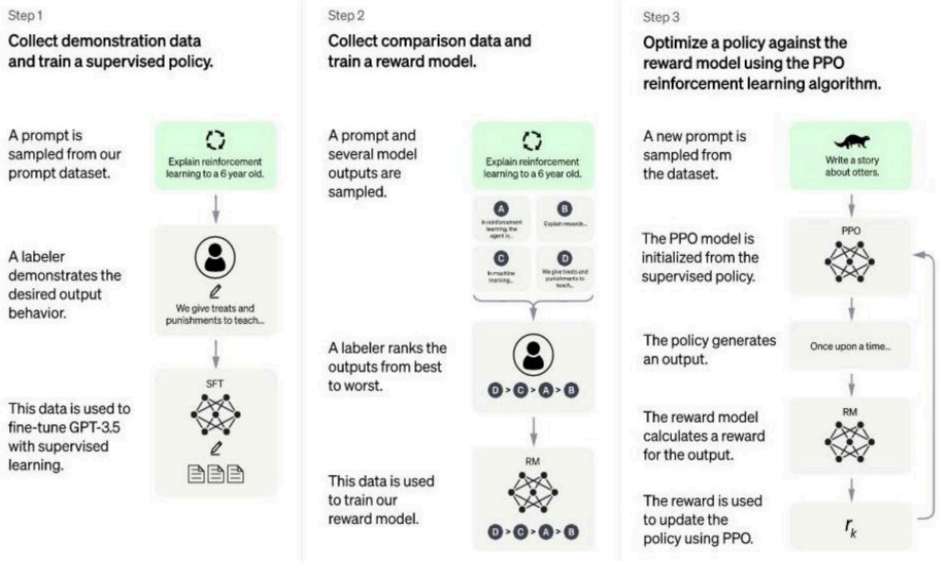
1. 데이터 수집 및 전처리 과정
2. 모델 선택 및 학습 방법론
3. 평가 및 최적화 단계
4. 추가 미세조정
5. NPU 버전으로 확장

공개되어 있는 다양한 말뭉치를 수집 후,
학습할 수 있는 형태로 전처리 수행

번호	폴더명	어절 수
1	WIKI pedia 영문	27억 어절
2	WIKI pedia 한글	0.8억 어절
3	초거대AI 헬스케어 질의응답 데이터	2억 어절
4	한국어 성능이 개선된 초거대 AI 말뭉치	20억 어절
5	의료, 법률 전문서적 말뭉치	0.5억 어절

LLM이 지시(Instruct)에 따르게 미세조정(Finetuning) RLHF
지도학습 기반 미세조정(SFT), 보상모델학습(RM), PPO 기반 강화학습

<InstructGPT 구축과정>



LLM의 성능을 평가하기 위한 benchmark

Task	Tested Concepts	Supercategory
Abstract Algebra	Groups, rings, fields, vector spaces, ...	STEM
Anatomy	Central nervous system, circulatory system, ...	STEM
Astronomy	Solar system, galaxies, asteroids, ...	STEM
Business Ethics	Corporate responsibility, stakeholders, regulation, ...	Other
Clinical Knowledge	Spot diagnosis, joints, abdominal examination, ...	Other
College Biology	Cellular structure, molecular biology, ecology, ...	STEM
College Chemistry	Analytical, organic, inorganic, physical, ...	STEM
College Computer Science	Algorithms, systems, graphs, recursion, ...	STEM
College Mathematics	Differential equations, real analysis, combinatorics, ...	STEM
College Medicine	Introductory biochemistry, sociology, reasoning, ...	Other
College Physics	Electromagnetism, thermodynamics, special relativity, ...	STEM
Computer Security	Cryptography, malware, side channels, fuzzing, ...	STEM
Conceptual Physics	Newton's laws, rotational motion, gravity, sound, ...	STEM
Econometrics	Volatility, long-run relationships, forecasting, ...	Social Sciences
Electrical Engineering	Circuits, power systems, electrical drives, ...	STEM
Elementary Mathematics	Word problems, multiplication, remainders, rounding, ...	STEM
Formal Logic	Propositions, predicate logic, first-order logic, ...	Humanities
Global Facts	Extreme poverty, literacy rates, life expectancy, ...	Other
High School Biology	Natural selection, heredity, cell cycle, Krebs cycle, ...	STEM
High School Chemistry	Chemical reactions, ions, acids and bases, ...	STEM
High School Computer Science	Arrays, conditionals, iteration, inheritance, ...	STEM
High School European History	Renaissance, reformation, industrialization, ...	Humanities
High School Geography	Population migration, rural land-use, urban processes, ...	Social Sciences
High School Gov't and Politics	Branches of government, civil liberties, political ideologies, ...	Social Sciences
High School Macroeconomics	Economic indicators, national income, international trade, ...	Social Sciences
High School Mathematics	Pre-algebra, algebra, trigonometry, calculus, ...	STEM
High School Microeconomics	Supply and demand, imperfect competition, market failure, ...	Social Sciences
High School Physics	Kinematics, energy, torque, fluid pressure, ...	STEM
High School Psychology	Behavior, personality, emotions, learning, ...	Social Sciences
High School Statistics	Random variables, sampling distributions, chi-square tests, ...	STEM
High School US History	Civil War, the Great Depression, The Great Society, ...	Humanities
High School World History	Ottoman empire, economic imperialism, World War I, ...	Humanities
Human Aging	Senescence, dementia, longevity, personality changes, ...	Other
Human Sexuality	Pregnancy, sexual differentiation, sexual orientation, ...	Social Sciences
International Law	Human rights, sovereignty, law of the sea, use of force, ...	Humanities
Jurisprudence	Natural law, classical legal positivism, legal realism, ...	Humanities
Logical Fallacies	No true Scotsman, base rate fallacy, composition fallacy, ...	Humanities
Machine Learning	SVMs, VC dimension, deep learning architectures, ...	STEM
Management	Organizing, communication, organizational structure, ...	Other
Marketing	Segmentation, pricing, market research, ...	Other
Medical Genetics	Genes and cancer, common chromosome disorders, ...	Other
Miscellaneous	Agriculture, Fermi estimation, pop culture, ...	Other
Moral Disputes	Freedom of speech, addiction, the death penalty, ...	Humanities
Moral Scenarios	Detecting physical violence, stealing, externalities, ...	Humanities
Nutrition	Metabolism, water-soluble vitamins, diabetes, ...	Other
Philosophy	Skepticism, phronesis, skepticism, Singer's Drowning Child, ...	Humanities
Prehistory	Neanderthals, Mesoamerica, extinction, stone tools, ...	Humanities
Professional Accounting	Auditing, reporting, regulation, valuation, ...	Other
Professional Law	Torts, criminal law, contracts, property, evidence, ...	Humanities
Professional Medicine	Diagnosis, pharmacotherapy, disease prevention, ...	Other
Professional Psychology	Diagnosis, biology and behavior, lifespan development, ...	Social Sciences
Public Relations	Media theory, crisis management, intelligence gathering, ...	Social Sciences
Security Studies	Environmental security, terrorism, weapons of mass destruction, ...	Social Sciences
Sociology	Socialization, cities and community, inequality and wealth, ...	Social Sciences
US Foreign Policy	Soft power, Cold War foreign policy, isolationism, ...	Social Sciences
Virology	Epidemiology, coronaviruses, retroviruses, herpesviruses, ...	Other
World Religions	Judaism, Christianity, Islam, Buddhism, Jainism, ...	Humanities

Table 2: Summary of all 57 tasks.

https://paperswithcode.com/dataset/mmlu

MMLU (Massive Multitask Language Understanding)는 모델을 제로샷 및 퓨샷 설정에서만 평가하여 사전 학습 중에 습득한 지식을 측정하도록 설계된 벤치마크
벤치마크는 STEM, 인문학, 사회 과학 등 57개 과목을 포괄하며, 초급 수준에서 고급 전문가 수준까지 난이도가 다양하며 세계적 지식과 문제 해결 능력을 모두 테스트함

번호	Task (영문)	(한글)	문항 수
1	anatomy	해부학	135
2	clinical_knowledge	임상지식	265
3	college_medicine	대학의학	173
4	medical_genetics	의학 유전학	100
5	professional_medicine	전문 의학	272

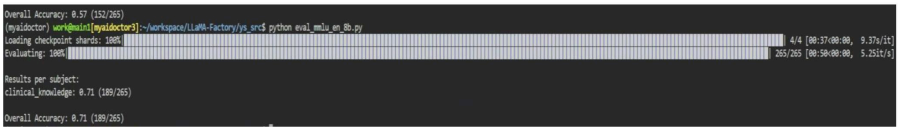
https://arxiv.org/pdf/2009.03300v3.pdf .

llama 3.1 8b instruct 기반 모델에 MMLU 중 의학관련 평가결과

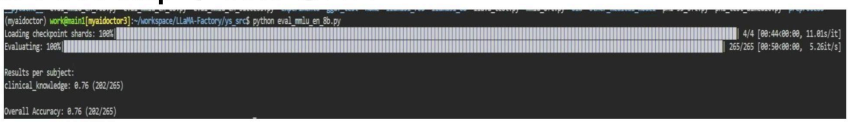
모델명	차수	결과	문자열 추론
llama31_8b_instruct	1	anatomy Accuracy: 0.67 (90/135) clinical_knowledge Accuracy: 0.74 (195/265) college_medicine Accuracy: 0.66 (114/173) medical_genetics Accuracy: 0.70 (70/100) professional_medicine Accuracy: 0.78 (213/272) All Accuracy Mean value: 0.71	공복 혈당 120 이상으로 나타난 경우에는 만약 당뇨병이 있는지 확인하는 데 도움이 됩니다. 제1형 당뇨 환자의 경우 일반적으로 인슐린 투여나 oral hypoglycemic agent(구토산성 수용체 작용제 등) 처방이 이루어진다. 또한 체중 조절 및 운동량 증가도 권고된다. 제2형 당뇨 환자의 경우에도 구토산성 수용체 작용제, GLP-1 가수분해 억제제, DPP-4억제제 등의 Орально 항당뇨제 또는 Insulin analogs가 사용될 수 있다. 또한 생활 습관 개선(식습관 변화 및 신체활동 증가)은 중요한 요소이다. 환자별 다양한 요인을 고려하여 개인 맞춤적인 치료 계획이 필요합니다.
llama318bik2b_ml3	1	anatomy Accuracy: 0.72 (97/135) clinical_knowledge Accuracy: 0.75 (198/265) college_medicine Accuracy: 0.67 (116/173) medical_genetics Accuracy: 0.71 (71/100) professional_medicine Accuracy: 0.76 (208/272) All Accuracy Mean value: 0.72	공복 혈당이 120 이상으로 나타난다는 것은 당뇨병 발병 위험성이 높아지는 신호일 수 있습니다. 따라서, 이를 확인하기 위해서는 혈당 조절을 위한 적절한 관리가 필요합니다. 만약 혈당 관리를 잘못하면 합병증 발생 가능성도 증가할 수 있으므로 주의해야 합니다. 제1형 당뇨환자의 경우에는 식사 후 혈당 조절을 위해 인슐린 사용이나 저용량 메틸프레드니솔론 투여 등의 약물을 고려할 수 있습니다. 반면에 제2형 당뇨병 환자들은 식습관 개선 및 체중 감소 등을 통해 혈당을 효과적으로 유지하는 것이 중요합니 다. 또한, 건강한 라이프스타일을 유지하여 당뇨의 진행 속도를 늦출 수도 있습니다. 따라서, 당뇨병을 앓고 있는 사람들에게는 꾸준한 혈당 모니터링과 개인적인 노력이 필수적이며, 정기적인 운동과 균형 잡힌 식사를 통해 혈당 조절을 도모해야 합니다. 특 히, 고혈압, 비만 등 다른 심각한 합병증이 동반될 때에는 전문의와 상담하여 추가적인 조치를 취하는 것이 좋습니다. 또한, 혈당 변동성을 최대한 줄이는 것도 중요한데, 이는 식사 시점을 규칙적으로 하고 규칙적인 운동을 실천함으로써 이루어진다거나, 스 트레스나 우울감 등을 피하며 충분한 휴식을 취하는 것 등이 도움이 될 수 있습니다. 마지막으로, 정기적인 검진을 받으면서 예방 접종을 실시하는 것도 매우 중요합니다. 의사의 지도 하에 꾸준한 관리와 감독을 받으며, 일상생활에서의 당류 섭취와 영양제 복용에도 주의를 기울여야 합니다. 이러한 방법들을 통해 당뇨병의 합병증을 예방하고 삶의 질을 향상시키는 데 큰 도움을 받을 수 있습니다. 그러나 당뇨병에 대한 이해도가 부족하다면 전문가의 조언을 듣지 않고 스스로 치료하지 않도록 주의해야 합니다.

mmlu의 clinical_knowledge 기준으로
SFT, DPO 데이터셋을 만들어 학습시킨 예

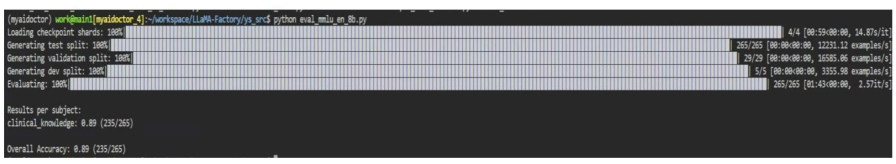
Full Finetuning 미세조정후 71% 정답률



SFT 500 step 후 76% 정답률



DPO 200 step 후 89% 정답률



사용 시 주의사항

❍ NPU는 허깅페이스에 있는 모든 AI 모델을 지원하는 것은 아님. 사용하고자 하는 모델을 해당 NPU에서 지원하는지 사전 확인해야 함

Decoder-only Models

Model Name	Architecture	Example Hugging Face Models
CodeLlama	`LlamaForCausalLM`	`meta-llama/CodeLlama-7b-Instruct-hf`
DeepSeek R1	`LlamaForCausalLM`	`deepseek-ai/DeepSeek-R1-Distill-Llama-8B`, `deepseek-ai/DeepSeek-R1-Distill-Llama-70B`
EXAONE 3.0, EXAONE 3.5	`ExaoneForCausalLM`	`LGAI-EXAONE/EXAONE-3.0-7.8B-Instruct`, `LGAI-EXAONE/EXAONE-3.5-7.8B-Instruct`
Llama 2, Llama 3.1, Llama 3.3	`LlamaForCausalLM`	`meta-llama/Llama-3.1-8B-Instruct`, `meta-llama/Llama-3.1-70B-Instruct`, `meta-llama/Llama-3.3-70B-Instruct`, ..
Solar	`LlamaForCausalLM`	`upstage/SOLAR-10.7B-v1.0`, `upstage/SOLAR-10.7B-Instruct-v1.0`
Qwen 2, Qwen 2.5	`Qwen2ForCausalLM`	`Qwen/Qwen2.5-Coder-32B-Instruct`, `Qwen/QwQ-32B`

https://developer.furiosa.ai/latest/en/overview/supported_models.html

사용 시 주의사항:

o 허깅페이스에 오픈되어 있는 모델 중 NPU에 맞게 변환된 모델을 사용해야 함. 아래는 furiosa-ai 에서 공개한 모델 중 일부임. https://huggingface.co/furiosa-ai

o 일반적으로 NVIDIA GPU에서 수행되는 모델을 바로 수행되지 않음. 아티팩트를 만드는 과정이 필요함

o NPU는 일반적으로 추론만 지원하고 학습은 지원하지 않음

Pre-compiled Model	Description	Base Model	Support Version
furiosa-ai/bert-large-uncased-INT8-MLPerf	INT8 quantized, optimized for MLPerf	google-bert/bert-large-uncased	2025.2
furiosa-ai/gpt-j-6b-FP8-MLPerf	FP8 quantized, optimized for MLPerf	EleutherAI/gpt-j-6b	2025.2
furiosa-ai/DeepSeek-R1-Distill-Llama-8B	BF16	deepseek-ai/DeepSeek-R1-Distill-Llama-8B	>= 2025.3
furiosa-ai/DeepSeek-R1-Distill-Llama-70B	BF16	deepseek-ai/DeepSeek-R1-Distill-Llama-70B	>= 2025.3
furiosa-ai/DeepSeek-R1-Distill-Qwen-7B	BF16	deepseek-ai/DeepSeek-R1-Distill-Qwen-7B	>= 2025.3

활용 방법

직접 구매하거나 CSP를 이용하여 해당 NPU를 사용할 수 있음

GPU | **NPU** | CPU

Instance	vRAM	CPU	RAM	On-demand
N-RAP-64 **4 x Rebellions ATOM+ 16GB**	64GB	24vCore	288GiB	₩10,400/hr
N-RAP-32 **2 x Rebellions ATOM+ 16GB**	32GB	12vCore	144GiB	₩5,200/hr
N-RAP-16 **Rebellions ATOM+ 16GB**	16GB	6vCore	72GiB	₩2,600/hr
N-FK-48 **FuriosaAI Renegade 48GB**	48GB	16vCore	192GiB	₩4,250/hr
N-FW-64 **4 x FuriosaAI Warboy 16GB**	64GB	16vCore	192GiB	₩5,600/hr
N-FW-32 **2 x FuriosaAI Warboy 16GB**	32GB	8vCore	96GiB	₩2,800/hr
N-FW-16 **FuriosaAI Warboy 16GB**	16GB	4vCore	48GiB	₩1,400/hr

https://elice.cloud/ko/pricing/elice-cloud

추론을 수행하게 되면 일반 GPU와 차이없음



2025년 과학기술정보통신부와 정보통신산업진흥원의 "고성능 컴퓨팅 지원(NPU)" 사업에서 "NPU 기반 sLLM 모델에서 헬스케어

공개된 모델의 직접 활용

furiosa-ai에서 공개한 모델

https://huggingface.co/furiosa-ai

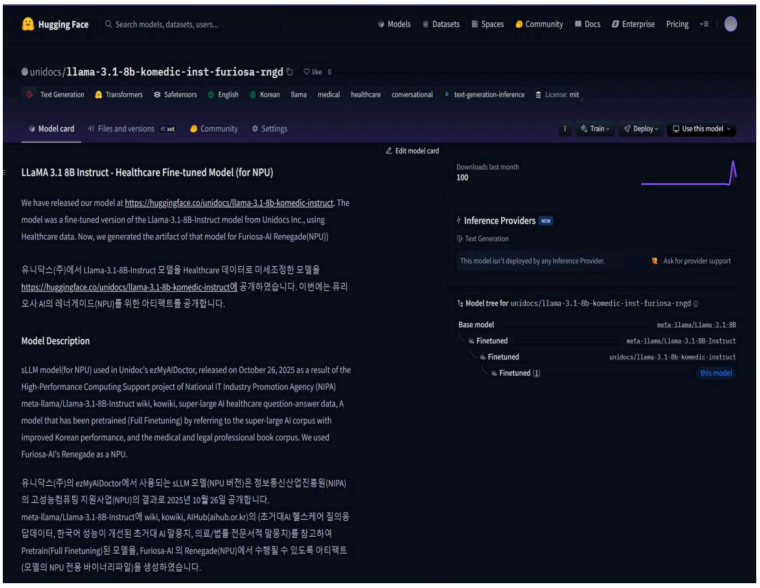
Pre-compiled models

Please check out the collection of models at https://huggingface.co/furiosa-ai/collections.

Pre-compiled Model	Description	Base Model	Support Version
furiosa-ai/bert-large-uncased-INT8-MLPerf	INT8 quantized, optimized for MLPerf	google-bert/bert-large-uncased	2025.2
furiosa-ai/gpt-j-6b-FP8-MLPerf	FP8 quantized, optimized for MLPerf	EleutherAI/gpt-j-6b	2025.2
furiosa-ai/DeepSeek-R1-Distill-Llama-8B	BF16	deepseek-ai/DeepSeek-R1-Distill-Llama-8B	>= 2025.3
furiosa-ai/DeepSeek-R1-Distill-Llama-70B	BF16	deepseek-ai/DeepSeek-R1-Distill-Llama-70B	>= 2025.3
furiosa-ai/DeepSeek-R1-Distill-Qwen-7B	BF16	deepseek-ai/DeepSeek-R1-Distill-Qwen-7B	>= 2025.3
furiosa-ai/DeepSeek-R1-Distill-Qwen-14B	BF16	deepseek-ai/DeepSeek-R1-Distill-Qwen-14B	>= 2025.3
furiosa-ai/DeepSeek-R1-Distill-Qwen-32B	BF16	deepseek-ai/DeepSeek-R1-Distill-Qwen-32B	>= 2025.3
furiosa-ai/EXAONE-3.5-7.8B-Instruct	BF16	LGAI-EXAONE/EXAONE-3.5-7.8B-Instruct	>= 2025.2

unidocs에서 공개한 모델

https://huggingface.co/unidocs/llama-3.1-8b-komedic-inst-furiosa-rngd



2025년 과학기술정보통신부와 정보통신산업진흥원의 "고성능 컴퓨팅 지원(NPU)" 사업에서 "NPU 기반 sLLM 모델에서 헬스케어



안전한 의료 AI를 위한 책임과 이슈

1. 인공지능 기본법
2. 하위 법령 및 가이드라인
3. 고영향AI 사업자책무 가이드라인

인공지능 기본법이 없을 때는 자의적 판단, 면책조항 고지

번호	구분	내용
1	개인정보 보호 및 데이터 보안	• 의료 데이터의 민감성과 개인정보보호법 준수 방안 • HIPAA(미국), GDPR(EU) 등 국제 개인정보보호 규정 준수 전략 • 데이터 비식별화 및 암호화 기술 적용 사례
2	의료 의사결정 지원 시스템으로서의 법적 책임	• AI 의사결정에 대한 책임 소재 문제 • 의료진과 AI 시스템 간의 역할 구분 및 법적 지위 • 오진 또는 의료사고 발생 시 책임 분배 방안
3	규제 준수 및 인증	• 의료기기로서의 AI 시스템 규제 현황 • FDA, CE 등 국제 인증 획득 절차 및 요구사항 • 국내 식품의약품안전처의 AI 의료기기 규제 대응 방안
4	윤리적 사용 및 편향성 문제	• AI 모델의 학습 데이터 편향성 문제와 법적 리스크 • 공정성, 투명성, 설명가능성 확보를 위한 기술적/제도적 장치 • AI 윤리 가이드라인 수립 및 준수 방안
5	지적재산권 및 라이선싱	• AI 모델 및 학습 데이터에 대한 지적재산권 이슈 • 오픈소스 활용에 따른 라이선스 준수 사항 • 특허 출원 및 보호 전략

Legal Disclaimer

The model developers and distributors bear no legal responsibility for any consequences arising from the use of this model.

This includes any direct, indirect, incidental, special, punitive, or consequential damages resulting from the model's output.

By using this model, users assume all risks that may arise, and the responsibility for verifying and appropriately using the model's output lies solely with the user.

This model cannot substitute for medical advice, diagnosis, or treatment, and qualified healthcare professionals should always be consulted for medical decisions.

This disclaimer applies to the maximum extent permitted by applicable law.

법적 책임 면책 조항

본 모델의 사용으로 인해 발생하는 모든 결과에 대해 모델 개발자 및 배포자는 어떠한 법적 책임도 지지 않습니다.

이는 모델의 출력으로 인한 직접적, 간접적, 우발적, 특수한, 징벌적 또는 결과적 손해를 포함합니다.

사용자는 본 모델을 사용함으로써 발생할 수 있는 모든 위험을 감수하며, 모델의 출력에 대한 검증 및 적절한 사용에 대한 책임은 전적으로 사용자에게 있습니다.

본 모델은 의학적 조언, 진단, 또는 치료를 대체할 수 없으며, 의료 관련 결정을 내릴 때는 반드시 자격을 갖춘 의료 전문가와 상담해야 합니다.

이 면책 조항은 관련 법률이 허용하는 최대 범위 내에서 적용됩니다.

AI 시대 자국 산업 발전과 AI로 인한 잠재적 위험에 대응하기 위해 자국 여건에 맞는 법제도 정립
- AI 혁신은 촉진, 위험은 최소화 - 미 AI위험관리 프레임워크

- AI 기본법 제정(24.12 국회통과, 26.1월 시행 예정)
- AI 기본법은 국내 AI 산업 현황을 고려하여 AI산업 발전과 인전 · 신뢰 기반 구축이라는 두 가치를 균형적으로 반영

AI기본법 주요 규정 사항

국가 AI 정책 거버넌스 확립	AI산업 육성 지원	안전신뢰 기반 조성
• 대통령을 위원장으로 하는 **국가AI위원회**를 설치하고, 개별 부처의 AI정책을 심의· 의결하도록 법률상 근거 마련 • 국가AI위원회 설치·기능, 분과·특별위원회, 지원단 설치 등	• 국가AI산업 발전을 위해, AI개발·도입·활용 전영역에서 **정부**가 다양한 지원을 할 수 있도록 **지원 사항** 마련 • AI R&D 지원, 기술 표준화, 학습용데이터 구축, AI기술 도입·활용지원, 법제도 개선 · 중소기업·창업 지원, 산업간 융합 촉진, 전문인력 확보, 국제협력, 해외진출 지원 등	• 안전하고 신뢰할 수 있는 AI 기반 조성을 통해 **AI의 활용을 촉진하고, 잠재적 위험으로부터 국민을 보호** • AI 윤리원칙, 민간자율 AI윤리위원회 설치 • 안전·신뢰 검·인증, 투명성·안전성 확보 의무, 고영향AI 확인·사업자 책무, AI 기본권 영향평가 등

AI기본법 하위법령 제정 방향, 2025, 과학기술정보통신부
https://nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=99835&bcIdx=28600&parentSeq=28600

헬스케어 sLLM 사업과 관련 인공지능 기본법의 내용 – 고영향 인공지능

제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

4. "**고영향 인공지능**"이란 **사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템**으로서 다음 각 목의 어느 하나의 영역에서 활용되는 것을 말한다.

다.「보건의료기본법」 제3조제1호에 따른 보건의료의 제공 및 이용체계의 구축・운영
라.「의료기기법」 제2조제1항에 따른 의료기기 및「디지털의료제품법」 제2조제2호에 따른 디지털의 료기기의 개발 및 이용
카. 그 밖에 사람의 생명・신체의 안전 및 기본권 보호에 중대한 영향을 미치는 영역으로서 대통령령으로 정하는 영역

8. "**이용자**"란 인공지능제품 또는 인공지능서비스를 제공받는 자를 말한다.
9. "**영향받는 자**"란 인공지능제품 또는 인공지능서비스에 의하여 자신의 생명, 신체의 안전 및 기본권에 중대한 영향을 받는 자를 말한다.

제3조(기본원칙 및 국가 등의 책무)

② 영향받는 자는 인공지능의 최종결과 도출에 활용된 주요 기준 및 원리 등에 대하여 기술적・합리적으로 가능한 범위에서 명확하고 의미 있는 설명을 제공받을 수 있어야 한다.

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 (인공지능기본법)

헬스케어 sLLM 사업과 관련 인공지능 기본법의 내용 – 투명성, 안전성

제31조(인공지능 **투명성 확보** 의무) ① 인공지능사업자는 **고영향 인공지능이나 생성형 인공지능을 이용한 제품 또는 서비스를 제공하려는 경우 제품 또는 서비스가 해당 인공지능에 기반하여 운용된다는 사실을 이용자에게 사전에 고지**하여야 한다.
② 인공지능사업자는 생성형 인공지능 또는 이를 이용한 제품 또는 서비스를 제공하는 경우 그 결과물이 생성형 인공지능에 의하여 생성되었다는 사실을 표시하여야 한다.
③ 인공지능사업자는 인공지능시스템을 이용하여 실제와 구분하기 어려운 가상의 음향, 이미지 또는 영상 등의 결과물을 제공하는 경우 해당 결과물이 인공지능시스템에 의하여 생성되었다는 사실을 이용자가 명확하게 인식할 수 있는 방식으로 고지 또는 표시하여야 한다. 이 경우 해당 결과물이 예술적·창의적 표현물에 해당하거나 그 일부를 구성하는 경우에는 전시 또는 향유 등을 저해하지 아니하는 방식으로 고지 또는 표시할 수 있다.

제32조(인공지능 **안전성 확보** 의무) ① 인공지능사업자는 학습에 사용된 **누적 연산량이 대통령령으로 정하는 기준 이상인 인공지능시스템의 안전성을 확보하기** 위하여 다음 각 호의 사항을 이행하여야 한다.
1. 인공지능 수명주기 전반에 걸친 **위험의 식별·평가 및 완화**
2. 인공지능 관련 **안전사고를 모니터링하고 대응하는 위험관리체계 구축**
② 인공지능사업자는 제1항 각 호에 따른 사항의 이행 결과를 과학기술정보통신부장관에게 제출하여야 한다.

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 (인공지능기본법)

헬스케어 sLLM 사업과 관련 인공지능 기본법의 내용 – 고영향인공지능 확인, 사업자 책무

제33조(**고영향 인공지능의 확인**) ① **인공지능사업자는 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우 그 인공지능이 고영향 인공지능에 해당하는지에 대하여 사전에 검토하여야 하며, 필요한 경우 과학기술정보통신부장관에게 고영향 인공지능에 해당하는지 여부의 확인을 요청할** 수 있다.
③ 과학기술정보통신부장관은 고영향 인공지능의 기준과 예시 등에 관한 가이드라인을 수립하여 보급할 수 있다.

제34조(고영향 인공지능과 관련한 **사업자의 책무**) ① 인공지능사업자는 고영향 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우 **고영향 인공지능의 안전성·신뢰성을 확보**하기 위하여 다음 각 호의 내용을 포함하는 조치를 대통령령으로 정하는 바에 따라 이행하여야 한다.
1. **위험관리방안의 수립·운영**
2. 기술적으로 가능한 범위에서의 인공지능이 도출한 최종결과, 인공지능의 최종결과 도출에 활용된 주요 기준, 인공지능의 개발·활용에 사용된 학습용데이터의 개요 등에 대한 설명 방안의 수립·시행
3. 이용자 보호 방안의 수립·운영
4. 고영향 인공지능에 대한 사람의 관리·감독
5. 안전성·신뢰성 확보를 위한 조치의 내용을 확인할 수 있는 문서의 작성과 보관
6. 그 밖에 고영향 인공지능의 안전성·신뢰성 확보를 위하여 위원회에서 심의·의결된 사항
② 과학기술정보통신부장관은 제1항 각 호에 따른 조치의 구체적인 사항을 정하여 고시하고, 인공지능사업자에게 이를 준수하도록 권고할 수 있다.

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 (인공지능기본법)

헬스케어 sLLM 사업과 관련 인공지능 기본법의 내용 - 영향평가

제35조(고영향 인공지능 **영향평가**) ① 인공지능사업자가 고영향 인공지능을 이용한 제품 또는 서비스를 제공하는 **경우 사전에 사람의 기본권에 미치는 영향을 평가(이하 "영향평가"라 한다)하기 위하여 노력**하여야 한다.
② 국가기관등이 고영향 인공지능을 이용한 제품 또는 서비스를 이용하려는 경우에는 영향평가를 실시한 제품 또는 서비스를 우선적으로 고려하여야 한다.
③ 그 밖에 영향평가의 구체적인 내용・방법 등에 관하여 필요한 사항은 대통령령으로 정한다.

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 (인공지능기본법)

인공지능 기본법 하위법령 및 가이드라인(안) 대국민 의견 수렴 수행중



AI기본법 하위법령 제정 방향, 2025, 과학기술정보통신부

26년 1월 AI기본법 시행을 위해 관련 하위법령 마련 추진

- 민간전문가 80여 명으로 구성된 '하위법령 정비단'을 구성·운영(1.15~)하며, 시행령·고시(2건)·가이드라인(5건)마련을 추진

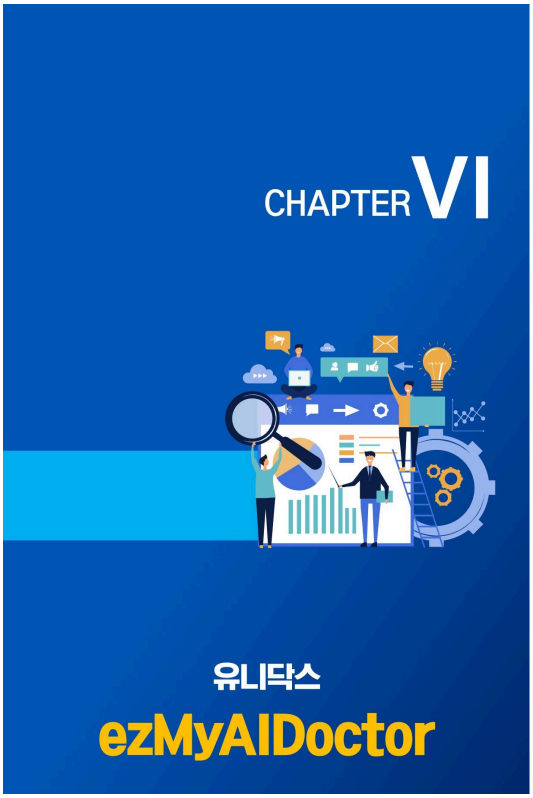
- ※ 하위법령 사항
  ❶시행령,
  ❷고영향AI 기준·예시 가이드,
  ❸고영향AI 사업자책무 고시 및 가이드
  ❹AI 안전성 확보 의무 고시 및 가이드
  ❺AI 투명성 확보 가이드
  ❻AI 영향평가 가이드

- 산업계·학계·시민단체·관계부처로 부터 총 70여차례 의견수렴을 토대로 하위법령초안을 작성. 국가AI전략위보고 및시행령 공개(9.8)

AI기본법 하위법령 제정 방향, 2025, 과학기술정보통신부
https://nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=99835&bcIdx=28600&parentSeq=28600

고영향AI에 대한 안전성/신뢰성 확보를 위한 사업자책무 조치방법과 책무 이행 간주 조항의 적용 범위를 가이드라인을 통해 구체화

구분	사업자 책무의 세부 내용(가이드라인)
위험관리방안의 수립·운영 (제1호)	▪ **위험관리정책 수립 및 이행** : 위험관리 담당 조직·인력을 중심으로 위험관리계획을 수립·이행하여, 고영향AI의 잠재적 위험을 식별·평가·제거해 허용가능한 수준으로 완화 ▪ **위험조직체계 수립 및 운영** : 위험 관리 담당 조직 구성 또는 조직 내 담당 인력 지정 ※ 사업자 규모·역량을 고려하여 겸임 인력 지정도 가능
설명 방안의 수립·시행 (제2호)	▪ **AI 도출 결과 및 주요 기준 근거 마련** : 고영향AI에 대한 설명방안 수립 시 가능 범위에서 투명성·설명가능성 제고를 위한 기술적 조치 필요 ▪ **학습용 데이터 정보의 관리** : AI에 활용된 학습용데이터의 일반적 내용과 함께 형식, 수량, 크기, 수집 및 전처리 방식 등 정보를 정의·관리 ▪ **설명 방안의 수립·시행** : 이용자에게 설명을 위한 절차, 범위, 방법 등을 검토하여 설명 방안을 수립·시행
이용자 보호방안 (제3호)	▪ **AI개발단계** : ▲적법·안전한 데이터 수집, ▲안전한 알고리즘 설계·시스템 개발 수행, ▲다양한 위험에 대비 충분한 평가 수행 ▪ **운영단계** : ▲문제시 실시간 모니터링 기반 마련, ▲이용자 피드백 프로세스 구축, ▲개인정보 등 이용자의 권리 보장을 위한 정책 수립
고영향AI에 대한 사람의 관리·감독 (제4호)	▪ **설계단계** : 고영향AI에 대해 사람이 AI의 동작에 개입할 수 있는 기준 및 절차·방법을 마련 ▪ **운영단계** : 성능저하·오류 예방을 위한 정기적 점검 계획·방안을 마련하고 개발사업자는 이용사업자에게, 이용사업자는 이용자에게 고영향AI의 이해도 향상을 위한 교육·훈련 방안 마련
문서의 작성·보관(제5호)	▪ 위 책무사항에 관한 **문서**(전자문서 포함) **작성** 및 **주기적 점검**, **최신 기술·방법론 적용**되도록 **관리**



Q&A

Track 1-5

Agentic AI 시대의 서비스 경험 혁신과 신뢰 거버넌스



김정훈
삼성전자 Principal Professional

발표개요

에이전틱 AI 시대에 AI가 가져올 서비스 경험 혁신과 효과 등을 공유하고, 나아가 AI 도입의 이면에 존재하는 고객 신뢰를 저해하는 리스크 요소들를 살펴보면서, 에이전틱 AI 시대에 부합하는 바람직한 '신뢰 거버넌스(Trust by Design)' 구축 방안 및 신뢰 원칙을 제시하며, 이를 글로벌 표준과 접목하고 실행하는 단계적 도입 로드맵 을 제안하고자 한다.

이력

- 삼성전자 글로벌CS센터 경영지원그룹
- 서비스 기획, 내부회계관리 및 IT 통제 PM
- ISACA Korea 아카데미 연구회 연구위원
- CISA, PMP, CPIM, IBM Cybersecurity Analyst
- 정보시스템 감리원, ISO27001, ISO9001 Auditor

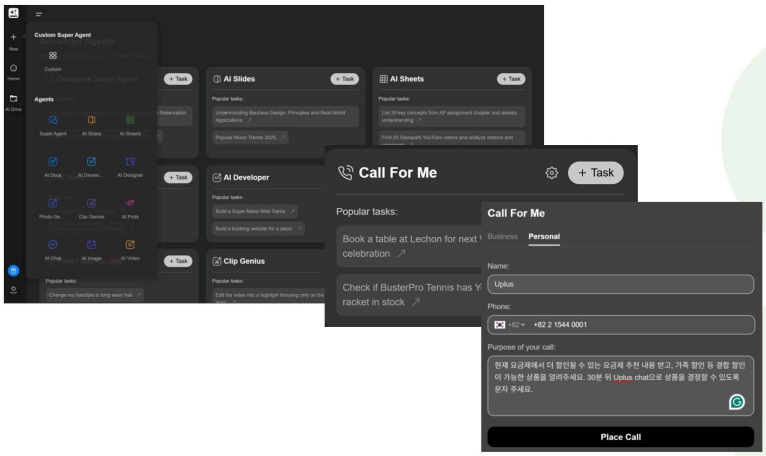
Agentic AI 시대의 서비스 경험 혁신과 신뢰 거버넌스

김정훈 I 삼성전자
ISACA 아카데미 연구회
www.linkedin.com/in/johnhoonkim

Confidential. For internal use only.

ISACA Korea Chapter

Agent AI 시대…



Agent to Agent Communication



두 고객의 이야기

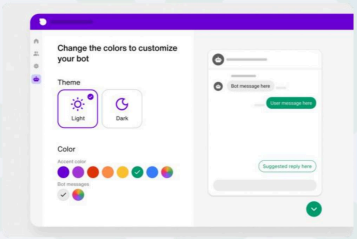
열 받는 경험



- 새 세탁기 문제로 고객센터에 전화했지만...
- 30분째 ARS와 씨름 중

"상담사 연결은 0번..." 무한 반복

깔끔한 고객 경험



- 문제 발생 전 카톡으로 알림 도착
- 배송 지연 사과와 함께 쿠폰 제공

"마치 제 마음을 읽는 집사 같은 서비스"

3

목차

1 **고객 경험**

무한 순환 여정과 AI 접목 영역

선제적 파트너십, 지능형 문제 해결, 고객 여정 오케스트레이션

2 **AI 도입 통한 고객 경험 혁신**

AI 기반 고객 경험에서의 신뢰 위협 요인 분석

고객 신뢰를 저해하는 6대 위험 요소와 영향 평가

3 **신뢰 거버넌스**

신뢰 기반 AI 고객 경험 구현을 위한 거버넌스 프레임워크

5대 원칙과 단계적 도입 전략

4

고객 경험의 무한 순환 여정
Infinite Customer Experience Journey Cycle

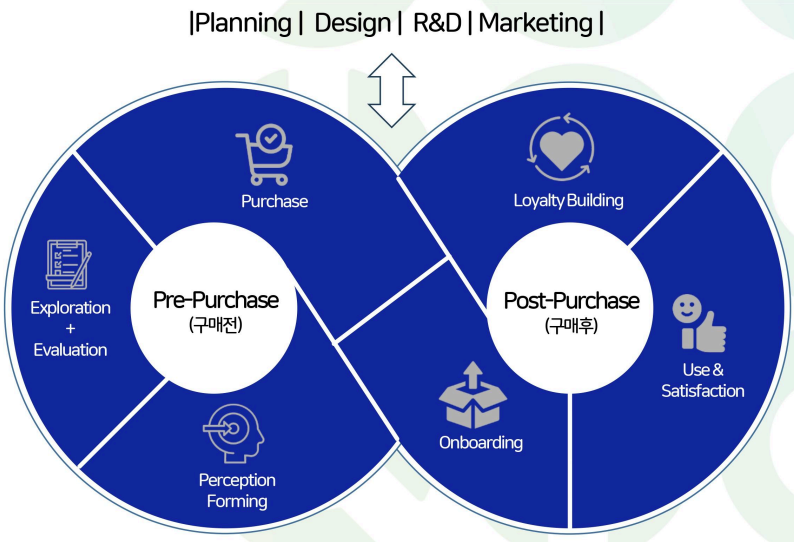
고객 경험의 중요성 증대

69%
미국 소비자는 Amazon 같은 디지털 고객 경험을 기대함

80%
미국 소비자는 제품만큼 고객 경험을 중요시 함

86%
미국 소비자는 더 나은 고객 경험에 비용을 지불할 의사가 있음

50%
미국 가정에는 최소 3개 이상의 삼성 제품을 보유하고 있음



[always-on experience with infinite pathways]

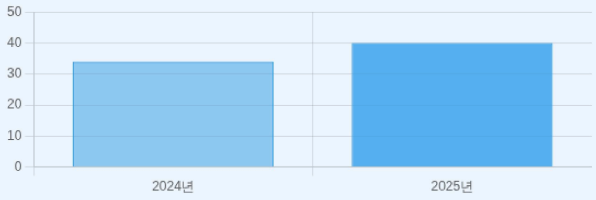
always-on experience with infinite pathways - 디지털과 물리적 접점을 넘나드는 무한한 고객 여정

에이전틱 AI 에 대한 기대 : 각 단계를 유기적으로 연결하며 개인화된 고객 경험 제공

5

Galaxy AI 경험

+6%p 18~29세 갤럭시 사용률 증가
한국갤럽 2025년 7월 조사 결과, 전년 대비



출처: 한국갤럽 조사, "스마트폰 관련 조사 2012-2025" (2025년 7월)

Galaxy AI 채택 현황

- Galaxy S25 사용자의 70%가 Galaxy AI 기능 활용
- 2025년 말까지 4억 대 기기에 Galaxy AI 제공 목표
- 사용자 경험과 접근성 지속 개선 중

출처: Samsung 공식 발표, "Galaxy AI adoption soars" (2025년 9월)

Generative Edit
AI로 사진 속 객체 자유롭게 편집, 이동, 생성

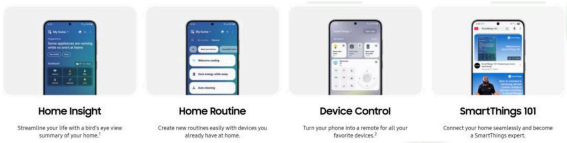
Circle to Search
화면 속 무엇이든 동그라미를 그려 즉시 검색

Live Translate
실시간 통화 및 메시지 번역, 13개국 언어 지원

Writing Assist
상황에 맞는 글쓰기 지원, 맞춤형 톤 조절

6

Multi-Device Experience



스마트폰을 넘어 냉장고, TV, 가전, 웨어러블이 모두 연결되는 AI 생태계(MDE, Multi-Device Experience)가 새로운 일상의 기준이 되고 있습니다.

글로벌 사용자 규모

3.5억

사용자

전 세계 SmartThings 생태계 사용자 수

출처: Samsung Global Newsroom, 2025

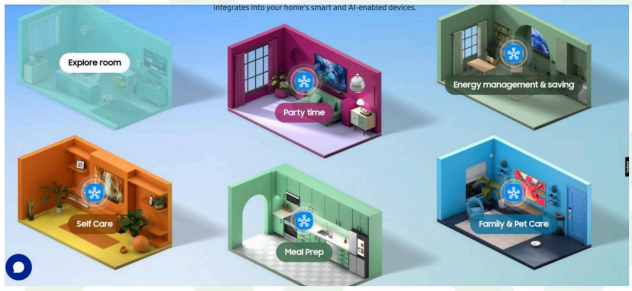
호환 기기 생태계

340

호환 브랜드

다양한 제조사와의 협력을 통한 확장된 기기 호환성

출처: Samsung Global Newsroom, 2025



Multi-Device Experience의 의미

통합 기기 연결

스마트폰, 냉장고, TV, 가전, 웨어러블 등 모든 기기가 하나의 생태계로 연결

AI 기반 지능화

인공지능이 사용자의 생활 패턴을 학습하고 맞춤형 경험 제공

에이전틱 AI로의 진화

단순 연결을 넘어 자율적 의사결정과 선제적 행동이 가능한 AI 경험

7

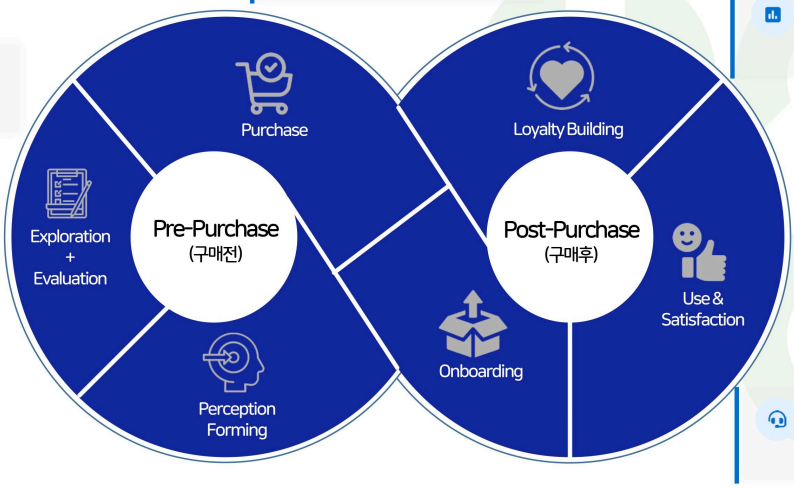
AI 접목 영역

Customer Experience Journey 단계별 IT 시스템 매핑

Reward System
Loyalty Building 단계
고객 충성도 보상 및 혜택 관리 시스템

Recommendation Tools
Pre-Purchase 단계
고객 관심사 기반 개인화 추천 시스템

Survey Tools
Post-Purchase 단계
고객 만족도 측정 및 피드백 수집 도구



Account / CRM
Purchase 단계
고객 계정 및 관계 관리 시스템

Customer Care
Post-Purchase 단계
고객 지원 및 문의 대응 시스템

IT 시스템 통합을 통한 끊김 없는 고객 경험 제공

8

경험 혁신 by AI

글로벌 기업의 AI 활용 고객경험 개선 사례

실제 사례 분석

Bank of America - Erica

- 20M 활성 사용자, 2.5B 고객 상호작용
- 98% 쿼리 해결률 달성
- 42M 이상의 고객 지원
- 고객 문의 자동화 및 개인화 금융 서비스

출처: Bank of America, 2024

Starbucks - Deep Brew AI

- 17M 앱 사용자, 전체 거래의 25% 처리
- AI 투자 대비 30% ROI 달성
- 개인화된 추천으로 판매 20% 증가
- 고객 경험 개인화 및 운영 효율화

출처: Forbes & The AI Report, 2025

Microsoft - Copilot

- 1,000+ 실제 기업 AI 도입 사례 보유
- 고객 서비스 응답 시간 60% 단축 (에너지 기업 사례)
- 고객 문의 해결률 35% 증가
- 다국어 지원 및 엔터프라이즈 통합 기능

출처: Microsoft, 2025

85%
고객 서비스 리더가 2025년 대화형 생성형 AI 도입 계획
Gartner 서베이, 2024

$83.9B
2033년 AI 고객 서비스 시장 규모 전망
25년 $15B
Grand View Research, 2024

65%
생성형 AI 정기적 사용 기업 비율
McKinsey Global Survey, 2024

"기업들은 AI 고객 서비스에 투자한 1달러당 평균 3.5달러의 수익을 올리고 있으며, 선도 기업들은 최대 8배의 ROI를 달성하고 있습니다. 그러나 64%의 고객들은 여전히 기업이 고객 서비스에 AI를 사용하지 않기를 원하고 있어, 신뢰 기반 접근이 필수적입니다."

- Gartner & Fullview AI Customer Service Report, 2025

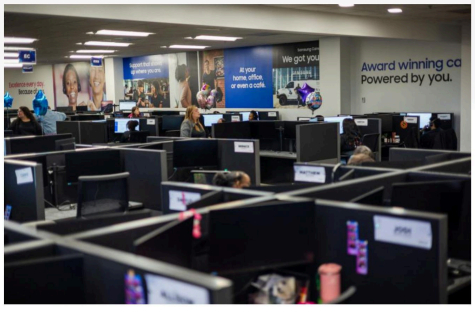
9

게임 체인저의 등장

기존 챗봇: 똑똑한 앵무새

- 정해진 질문에만 정해진 답변 제공
- 단일 질의-응답 패턴만 처리 가능
- 맥락 이해와 기억 능력 제한적
- 사용자가 질문한 내용에만 수동적 대응

78% 기업이 AI를 최소 한 개 이상의 비즈니스 기능에 활용 중
McKinsey Global Survey on AI, 2024



에이전틱 AI: 생각하고 행동하는 주체

- 목표를 이해하고 자율적으로 계획을 세움
- 여러 앱과 시스템을 통합 조정하며 실행
- 문제 해결을 위한 능동적 의사결정 및 실행
- 고객 경험 속 진정한 '행동하는 주체'로 등장

30% 2025년까지 기업이 AI 기반 여정 오케스트레이션을 도입 예정
IDC FutureScape Worldwide AI and Automation 2024

80% 2029년까지 에이전틱 AI가 고객 서비스 이슈의 80%를 자동 해결 예상
Gartner Predictions, 2025

10

Vision 1. 선제적 파트너

AI 도입 현황: McKinsey Global Survey 2024

출처 McKinsey Global Survey on AI, 2024

65%
조직이 생성형 AI를 정기적으로 사용(전년 대비 2배 증가)

78%
AI를 최소 한 개 이상의 비즈니스 기능에 활용
(전년 72%에서 증가)

53%
C레벨 임원이 생성형 AI 사용 주도(가장 높은 도입 그룹)

선제적 AI 파트너의 작동 원리 (How It Works) 검증된 메커니즘

검증됨

1. 예측 알고리즘

과거 데이터, 사용자 행동 패턴, 환경 변수를 분석하여 미래 니즈 예측

- 머신러닝 학습 모델
- 패턴 인식 및 상관관계 분석
- 지속적 모델 업데이트

검증됨

2. 실시간 데이터 분석

현재 상황과 예측 모델을 통합하여 최적의 대응 시점과 방법 결정

- 센서 및 API 연동
- 맥락 인식 분석
- 임계치 기반 트리거

검증됨

3. 프로액티브 액션

자율적으로 적절한 시기에 선제적 조치를 취하거나 제안 제공

- 맞춤형 알림 및 제안
- 자동 문제 해결
- 지속적 피드백 학습

Vision 1. 선제적 파트너

글로벌 기업의 선제적 AI 파트너 구현 방법론

Microsoft Copilot

고객 문의 처리 시간 40% 단축

구현 방법론:
- 프로액티브 프롬프트 생성 기술
- 대화 맥락 분석으로 자동 제안
- 15억 개 이상 데이터 포인트 분석

출처: Microsoft, "Proactive prompts and insights for Copilot", 2024

Bank of America Erica

27억+ 고객 상호작용, 98% 쿼리 처리율

구현 방법론:
- 예측 분석 알고리즘(패턴 탐지)
- 선제적 재무 경고 시스템
- 고객 행동 기반 맞춤형 제안

출처: Bank of America, "Erica's Proactive Capabilities", 2025

선제적 AI 핵심 메커니즘 (IBM Research, 2024)

선제적 AI 파트너는 3단계 메커니즘으로 작동: 1) 예측 알고리즘으로 고객 요구 예측, 2) 실시간 데이터 분석으로 최적 대응 결정, 3) 자율적 프로액티브 액션 제공. IBM 연구에 따르면 이 방식으로 고객 서비스 효율성 24% 향상, 선제적 문제 해결률 42% 증가

출처: IBM, "AI in Customer Service", 2024

Gartner 예측: 2029년까지 에이전틱 AI가 일반적인 고객 서비스 문제의 80%를 자율적으로 해결하여 운영 비용 30% 절감 (Gartner, 2025)

Vision 2. 궁극의 문제 해결사

"새로 산 세탁기가 갑자기 클럽 음악 같은 굉음을 내며 멈췄습니다."

실제 사례: 삼성 Home Appliance Remote Management

1. 문제 등록
SmartThings 앱으로 등록

2. 원격 진단
에이전틱 AI가 데이터 분석

3. 실시간 지원
화면 공유로 해결책 안내

4. 방문 최소화
원격 해결 → 만족도 상승

실제 원격 해결 사례 (검증 완료)

- 버튼 무반응 문제: Child Lock 설정 원격 확인 및 해제 안내
- 냉장고 도어 결로 현상: 원격으로 내부 히터 활성화 해결
- 화면 관련 문제: 화면 공유로 앱 오작동 진단
- 방문 수리 효율화: 사전 진단으로 적절한 부품과 도구 준비

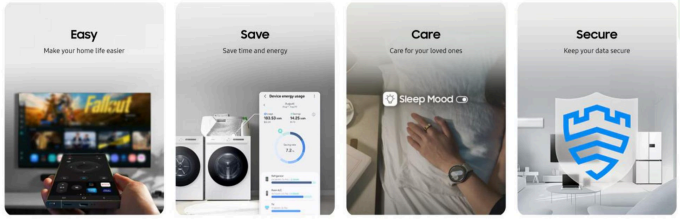
출처: Samsung Electronics 공식 발표 (2025년 8월)

AI 기반 고객 서비스 도입 효과 (2025) (2025 데이터)

- AI 도입 기업의 85%
  고객 만족도 평균 18% 향상
- 고객 문의 처리 시간 42% 단축
  반복적인 문의 처리 자동화로 효율성 증가
- 상담사 생산성 33% 향상
  복잡한 문제에 집중할 수 있는 환경 조성
- 에이전틱 AI 도입 27% 증가
  단순 챗봇에서 자율 의사결정 시스템으로 전환

출처: Deloitte Customer Service Excellence 2025 & Gartner 2025 Customer Service AI Use Cases

검증 기관: Gartner & Deloitte 2025



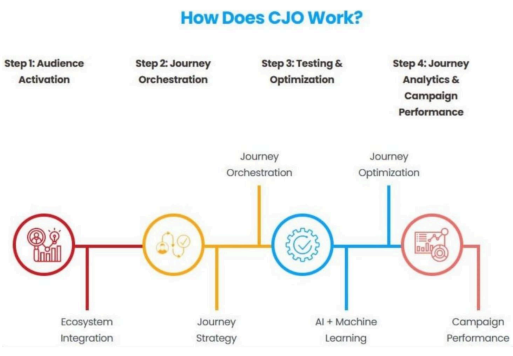
13

Vision 3. 고객 여정의 오케스트레이션

고객 여정 전체를 지휘하는 AI 오케스트레이터

"2025년까지 30%의 기업이 고객 경험과 참여를 개선하기 위해 AI 기반 여정 오케스트레이션 도입 예정" - IDC FutureScape: 글로벌 AI 예측 2024

How does CJO work? 출처: IDC 2024



AI 여정 오케스트레이션 5단계 프로세스

1. 문제 예측 — 사용 패턴 분석
2. 통합 분석 — 데이터 통합
3. 지능형 라우팅 — 최적 채널 선택
4. 자동 해결 — AI 문제 해결
5. 경험 연속성 — 옴니채널 일관성

검증된 AI 오케스트레이션 사례

- BOA Bank of America Erica (검증 완료)
  - 27억+ 고객 상호작용
  - 98% 쿼리 처리율
  - Bank of America 공식 발표 자료 기반
- MS Microsoft Copilot (검증 완료)
  - 1,000+ 기업 적용
  - 상담시간 38% 단축
  - Microsoft 공식 발표 자료 기반

AI 오케스트레이션 도입 성과

- 문제 해결 시간: 42% 단축
- 운영 비용: 25% 절감
- 고객 유지율: 20% 증가

14

IDC Worldwide Future of Customer Experience 2024 Predictions

신뢰의 플라이휠

AI 고객 경험의 선순환 구조



AI 고객 경험 실제 효과

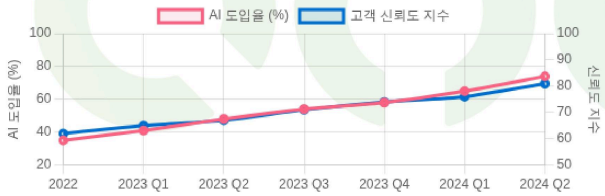
고객 서비스 ROI
+213%
AI 지원 상담 도입 기업
출처: Salesforce, Wiley 사례 (2024)

AI 챗봇 활용 계획
85%
고객 서비스 리더들의 2025년 도입 계획
출처: Gartner 서베이 (2024)

고객 생애 가치(CLV)
+65%
AI 기반 신뢰 생태계 구축 기업
출처: Salesforce 연구 (2024)

투자 대비 수익
$3.5
AI 고객 서비스에 투자한 $1당 수익
출처: Accenture 리포트 (2024)

AI 채택과 고객 신뢰도 관계



출처: McKinsey Global Survey on AI (2024) & Gartner Customer Trust Index (2024)

15

양날의 검

에이전틱 AI의 멋진 비전은, 동전의 양면처럼 어두운 면도 가지고 있습니다. Deloitte의 2024년 조사에 따르면, 기업의 65%가 AI 도입 시 고객 신뢰를 최우선 고려 사항으로 꼽았으며, 53%는 윤리적 위험 및 평판 손상을 우려한다고 답했습니다.

출처: Deloitte Global AI Survey, 2024

혁신 ⟷ 위험

혁신의 가능성

- 고객 서비스 비용 35% 감소 (Accenture Technology Vision 2025)
- 85%의 고객 서비스 리더가 2025년까지 대고객 생성형 AI 도입 계획 (Gartner, 2025)
- AI 활용 기업의 78%가 최소 하나 이상의 비즈니스 기능에서 성과 개선 (McKinsey State of AI, 2025)
- AI 고객 서비스에서 4배 ROI 달성 (Deloitte Customer Service Excellence, 2025)

신뢰 붕괴의 위험

- 고객의 최대 우려사항: 인간적 소통 부재 (Gartner, 2025)
- 소비자의 70%가 데이터 프라이버시와 보안에 대해 우려 (Deloitte Connected Consumer, 2025)
- 서비스/지원 리더의 77%가 AI 배치에 대한 경영진 압박 경험 (Gartner, 2025)
- 신뢰 회복에 평균 21개월 소요 (PwC 2025 Customer Experience Survey)

16

Deloitte TrustID™ 2023 조사: "AI 도구를 제공하는 기업에 대해 직원들은 188% 더 낮은 공감성과 인간미를 느낍니다."

신뢰를 무너뜨리는 6대 리스크

Gartner : 64%의 고객은 기업이 고객 서비스에 AI를 사용하지 않기를 원함

"개인은 AI를 활용하는 경험을 즐겨하면서, 서비스나 경험은 Human을 더 선호하는 경향"

공감 결여 48%

"Deloitte(2025): 고객의 53%가 AI와의 소통에서 인간적 공감 부족으로 부정적 경험 보고"

과도한 개입 61%

"Deloitte(2025): 61%의 소비자가 AI 서비스가 과도하게 개인화되었을 때 불편함을 느낌"

자동화의 덫 85%

"Gartner(2025): 85%의 고객 서비스 리더가 Gen AI 솔루션 탐색 계획, 하지만 인간 접점 부족 우려"

책임 전가의 악몽 58%

""PwC(2025): 58%의 소비자가 AI 도구와 브랜드 상호작용에 불편함 느낌"

데이터 불투명성 73%

"Accenture(2024): 73%의 고객이 기업이 자신의 데이터를 어떻게 사용하는지 투명하게 공개하길 원함"

편향된 의사결정 NIST

""NIST AI RMF 2025 Release 5.2.0: AI 시스템의 편향성은 최우선 위험요소로 지속적 모니터링 필요"

17

"이 중 하나라도 경험한 고객의 68%가 해당 브랜드와의 관계를 재고한다."

출처: Gartner(2024), Deloitte(2025), Accenture(2024), NIST(2024)

신뢰는 복구 비용이 크다



""소비자의 58%는 브랜드와의 상호작용에 AI 도구를 사용하는 것에 대해 약간 또는 전혀 편안함을 느끼지 못합니다."- PwC 2025 Customer Experience Survey

신뢰 회복의 실제 비용

- 신뢰 회복 비용: 신뢰 구축 비용의 5-7배 (Deloitte TrustID Survey, 2024)
- 고객 재획득 비용: 신규 고객 비용의 25배 (Accenture CX Report, 2024)
- 브랜드 이미지 회복 기간: 평균 2-3년 (Accenture "Customer service on the brink")
- AI 불신으로 인한 고객 이탈률: 53%(Gartner, 5,728명 고객 조사, 2024)

Deloitte의 3단계 신뢰 회복 모델

1. 신뢰 침해 확인 (24시간 이내)
   투명한 소통으로 AI 오류와 영향 인정 및 즉각 공개
2. 능동적 복구 (1-6개월)
   프로세스 재설계, 고객 컨트롤 강화 및 지속적인 소통 유지
3. 신뢰 재구축 (6개월-3년)
   일관된 가치 제공과 고객 중심 설계로 강화된 신뢰 구축

출처: Accenture Life Trends (2025), PwC Customer Experience Survey (2025), Deloitte Connected Consumer Survey (2025)

18

신뢰를 설계하라

Trust by Design

해답은 더 똑똑한 기술이 아니라, 더 견고한 거버넌스입니다.

우리가 재무 정보의 신뢰성을 위해 내부통제(COSO)를 도입하듯, AI 고객 경험에도 처음부터 '신뢰'를 의도적으로 설계하고 심어야 합니다.

기술 중심 접근	신뢰 중심 접근
"더 똑똑한 AI"	"더 견고한 거버넌스"

Trust by Design 프레임워크

1. 신뢰 원칙 정의
   명확한 윤리적 지침과 고객 중심 가치 설정
2. 거버넌스 구조 구축
   역할, 책임, 검증 체계 수립
3. 개발 프로세스 통합
   기술 설계 단계부터 신뢰 요소 내재화
4. 지속적 모니터링
   신뢰 지표 추적 및 개선

COBIT, NIST AI RMF 같은 글로벌 표준 프레임워크는 모두 '책임성', '투명성', '공정성'을 강조합니다.

19

원칙 1. 인간 중심 탈출구

NIST AI RMF 2025 가이드라인 NIST AI RMF 5.2.0

NIST AI RMF 5.2.0 (2025년 8월 릴리스)는 "강화된 인간 중심 AI 협력(Enhanced Human-AI Collaboration)"을 핵심 원칙으로 제시합니다:

모든 에이전틱 AI 시스템에 '역할 기반 인간 개입 메커니즘(Role-Based Human Intervention)' 의무화 (NIST.AI.600-2.5)

자율 의사결정 편향을 방지하기 위한 투명한 의사결정 기록 및 설명 제공 필수화 (Executive Order 14306)

McKinsey 2025 보고서: 인간-AI 하이브리드 협업 모델이 순수 AI 모델보다 고객 만족도 42% 향상 입증

NIST 인간 중심 설계 원칙

- 명확성: 고객이 AI와 대화 중임을 인지
- 접근성: 3클릭 이내 인간 상담사 연결
- 컨텍스트 유지: 대화 내용 자동 전달
- 투명성: 전환 시간과 이유 명확히 설명

Bank of America Erica AI 사례 연구

"복잡한 금융 요청 시 인간 상담사 연결 설계로 고객 신뢰도 향상"

- 도입 전: AI 단독 응대 시 복잡한 금융 문의 처리 불가 (고객 불만 54% 증가)
- 도입 후: 자연스러운 에스컬레이션 설계로 98% 쿼리 처리율 달성 (2024)

Gartner 2024 조사 결과

64%의 고객이 AI 없는 서비스 선호 vs
77%의 서비스 리더가 AI 도입 압박 받음

고객 우려 1위: 상담사 연결 어려움 (53%)

20

원칙 2. Brand Alignment

Deloitte 2024 연구결과: 고객의 77%가 브랜드 가치와 일치하지 않는 AI 응답에 부정적 반응을 보임
Accenture 소비자 조사(2025): 응답자의 81%는 브랜드의 디지털 경험이 일관되지 않을 때 신뢰도가 감소한다고 응답

스타벅스

브랜드 인격: 친근함, 지속가능성, 커뮤니티

Deep Brew AI - 실제 사례:

"오늘 비가 오는데 추천 메뉴가 있을까요?"

"비 오는 날에는 따뜻한 카라멜 마키아또가 어떨까요? 저희 매장에서 사용하는 우유는 지속가능한 낙농업을 지원하고 있어요. 더 풍부한 맛을 원하시면 에스프레소 샷을 추가하실 수 있어요."

출처: 스타벅스의 AI 기반 개인화 전략 (Forbes, 2025)

뱅크 오브 아메리카

브랜드 인격: 전문성, 신뢰성, 안정감

Erica AI 챗봇 - 실제 사례:

"최근 지출이 많았는데 어떻게 절약할 수 있을까요?"

"지난 3개월 데이터를 분석한 결과, 식비 지출이 15% 증가했네요. 더 효율적인 예산 관리를 위해 맞춤형 절약 계획을 제안해 드릴까요? 또한 당행의 Better Money Habits® 프로그램에서 추가 재정 관리 팁을 확인하실 수 있습니다."

출처: Bank of America의 Erica 27억 고객 상호작용 데이터 (2024)

연구 결과:
McKinsey(2024)에 따르면, 브랜드 톤과 일치하는 AI 상호작용은 고객 유지율을 23% 향상시키고 고객 평생 가치(CLV)를 28% 증가시킵니다.

21

AI는 단순한 도구가 아닌, 브랜드의 디지털 대변인 (Gartner, 2024)

원칙 3. 고객에게 투명성과 제어권 제공

고객에게 투명성과 제어권을 제공하라

EU AI Act(2024) 및 GDPR의 핵심 규제 요구사항: AI 시스템은 의사결정 과정을 설명할 수 있어야 하며, 데이터 활용에 대한 통제권을 제공해야 합니다.

- EU AI Act Article 13 - 고위험 AI 시스템은 투명성을 보장하고 결정 과정을 설명할 수 있어야 함
- GDPR Article 22 - 자동화된 의사결정에 대한 설명 요청 및 이의제기 권리 보장
- Accenture 연구(2024) - 소비자의 83%가 AI가 자신의 데이터를 어떻게 사용하는지 알 권리를 요구

실제 투명성 사례: 뱅크 오브 아메리카 Erica

Erica AI 챗봇은 27억 건 이상의 고객 상호작용에서 모든 금융 추천에 대해 " 추천 이유" 버튼을 제공, 투명성으로 98% 쿼리 처리율 달성 (2024 Fortune)

제어권 구현: GDPR 준수 사례

Deloitte의 2024 연구에 따르면 AI 설정 및 통제 기능을 제공하는 기업의 고객 신뢰도가 평균 42% 더 높음. 데이터 제어권이 신뢰의 핵심요소.

> *""투명성은 에이전틱 AI 시대에 선택이 아닌 필수입니다. Accenture Tech Vision 2025에 따르면 고객 경험에서 AI 의사결정의 투명성을 보장하는 기업은 고객 유지율이 31% 더 높고 신규 고객 유치 비용이 27% 낮습니다." - Accenture Life Trends 2025 Report*

22

원칙 4. 선제적 책임

Accenture(2024)의 "Customer Service on the Brink" 보고서에 따르면, 선제적 문제 해결 접근법을 도입한 기업의 고객 만족도는 평균 27% 증가하였습니다.

반응적 접근

- 고객이 문제를 발견하고 불만 제기
- 기업은 사후 대응 및 해명
- Gartner(2024): 65%의 고객이 이 단계에서 브랜드 전환 고려
- 고객 신뢰 회복에 평균 18개월 소요 (McKinsey, 2024)

선제적 접근

- AI가 자체 모니터링으로 오류 감지
- Deloitte(2024): 선제적 접근 도입 시 고객 유지율 34% 증가
- 적절한 보상으로 고객 만족도 42% 향상 (Accenture, 2024)
- 고객의 불필요한 시간 낭비 평균 76분 절약

실제 기업 적용 사례

> Bank of America의 AI 비서 'Erica'는 고객이 발견하기 전에 잠재적 거래 오류를 감지하고 즉시 해결책을 제시합니다. 2024년 1분기 기준 27억 건의 고객 상호작용 중 98%의 문제를 선제적으로 해결하여 고객 만족도 42% 향상 (Bank of America, 2024)

81%	45%	38%
선제적 대응 시 고객 신뢰도 증가	고객 만족도 향상 효과	고객 유지율 증가
McKinsey, 2025	Accenture, 2025	McKinsey, 2025

핵심 포인트: "에이전틱 AI가 문제를 먼저 발견하고 솔루션을 제시할 때 가장 신뢰받는다"

Accenture Technology Vision 2025: "선제적 책임 원칙을 적용한 기업은 고객 생애 가치(CLV)가 평균 34% 증가했으며, 77%의 경영진이 신뢰를 AI의 핵심 요소로 인식합니다."

23

원칙 5. 고객 중심 접근

AI에게 '고객에게 가장 좋은 것'과 '회사에 가장 좋은 것' 사이에서 항상 고객을 우선하도록 가르쳐야 합니다.

NIST AI RMF(2024)는 "AI 시스템은 인간 중심적이고 인간의 선택과 자율성을 보호하는 방향으로 설계되어야 한다"고 명시합니다.

기업 중심 접근법의 위험성

- Deloitte(2024): 54%의 소비자는 기업이 AI를 활용해 이익을 최대화할 것이라고 우려
- McKinsey(2024): AI 윤리 원칙 없는 기업의 고객 이탈률 32% 증가
- Gartner(2024): 64%의 소비자가 AI 고객 서비스 도입에 거부감 표시

EU AI Act 윤리적 가이드라인

- 투명성: 고객이 AI와 상호작용 중임을 인지할 권리
- 인간 감독: 모든 AI 결정에 인간의 감독 필요
- 개인정보보호: 데이터 사용에 대한 명시적 동의
- 편향 방지: 차별적 결정 방지를 위한 지속적 모니터링

실제 기업 사례: 윤리적 AI 결과

- Bank of America Erica AI(2025): 고객 금융 상태에 따라 더 저렴한 상품을 추천하여 고객 신뢰도 42% 향상, 충성도 37% 증가
- Microsoft Copilot(2025): 사용자에게 불필요한 기능 추천을 금지하는 윤리 가이드라인 강화 후 장기 구독률 29% 증가
- Accenture Technology Vision 2025: 고객 우선 윤리 원칙 도입 기업의 신뢰 기반 고객 관계 45% 향상, CLV(고객생애가치) 36% 증가

> *"AI 자율성이 확대되는 시대에, 윤리적 경계를 명확히 하는 기업만이 장기적 신뢰 자산을 구축할 수 있다." - Deloitte, Trust in AI Survey 2025*

24

CX/CS Governance

글로벌 표준과 5대 원칙의 통합 거버넌스 프레임워크

글로벌 표준 프레임워크

- **NIST AI RMF**
  위험관리 • 안전성 • 투명성
- **ISACA AI 거버넌스**
  감사 • 책임성 • 모니터링
- **ISO/IEC 42001**
  AI 관리 시스템 • 프로세스
- **EU AI Act**
  위험 기반 접근 • 규제 준수

CX/CS 신뢰 5대 원칙

1. **인간 중심 탈출구**
   AI 시스템의 한계 상황에서 인간 개입 보장
   NIST | EU Act
2. **Brand Alignment**
   AI가 브랜드 가치와 일관된 경험 제공
   ISACA
3. **고객에게 투명성과 제어권 제공**
   고객에게 의사결정 과정 투명성 제공
   NIST | ISO
4. **선제적 책임**
   AI 실수 발생 전 선제적 대응 체계
   ISACA | NIST
5. **고객 중심 접근**
   고객 이익을 최우선으로 하는 의사결정
   EU Act | ISO

신뢰 거버넌스 통합 체계

"85%의 고객 서비스 리더가 2025년에 생성형 AI 솔루션을 도입 예정이나, 64%의 고객은 AI 사용을 원치 않는 모순된 상황에서 신뢰 거버넌스가 핵심입니다."

- Gartner 고객 서비스 서베이 (2024)

글로벌 기업 적용 사례

- **Bank of America - Erica**
  AI 챗봇 27억 건 상호작용, 98% 쿼리 처리율
- **Microsoft Copilot**
  1,000개 이상 실제 기업 적용 성공
- **Starbucks**
  AI 예측 분석으로 30% ROI 달성
- **글로벌 기업 로드맵**
  McKinsey: 65%가 생성형 AI 정기 활용 (2024)

표준-원칙 연계 매트릭스

글로벌 표준	핵심 가치	연계 원칙
NIST AI RMF	투명성, 위험관리	원칙 1, 3, 4
ISACA AI 감사	책임성, 거버넌스	원칙 2, 4
ISO/IEC 42001	관리 체계, 프로세스	원칙 3, 5
EU AI Act	위험 기반 접근	원칙 1, 5

25

도입 로드맵 (Crawl-Walk-Run)

신뢰 중심의 단계적 AI 도입 전략



Crawl
기반 구축

Walk
확장 도입

Run
완전 통합

Gartner(2024): 고객 서비스 리더의 85%가 2025년까지 고객 대면 생성형 AI 솔루션을 탐색 또는 파일럿 예정

기반 구축 단계

내부 AI 어시스턴트 도입으로 상담사 생산성 향상

기존 시스템과의 통합 및 데이터 품질 확보

초기 ROI 측정 및 내부 피드백 체계 구축

Gartner 기준 성과: 응대시간 40% 단축

출처: Gartner Customer Service AI(2025)

확장 도입 단계

IDC 권장: 특정 고객 여정에 AI 시범 도입

인간-AI 협업 모델 및 에스컬레이션 경로 구축

고객 피드백 기반 개선 및 윤리적 검증

IDC 기준 성과: 자동화율 53% 달성

출처: IDC Artificial Intelligence Predictions(2024)

완전 통합 단계

Gartner 예측: 2029년까지 에이전틱 AI가 고객 서비스 이슈의 80% 자동 해결

선제적 고객 서비스 및 개인화된 경험 전면화

신뢰 거버넌스의 지속적 개선 및 준수 모니터링

성과 목표: 운영비용 30% 감소

출처: Gartner Predicts Agentic AI(2025)

26

McKinsey(2024): 조직의 65%가 정기적으로 생성형 AI를 사용 중, 그러나 도입은 단계적 접근이 필수적

핵심 요약

1

AI 혁신 역설

78%의 기업들이 AI를 최소 1개 비즈니스 기능에 활용하면서도, 70%의 소비자는 데이터 프라이버시 우려를 표현합니다.

출처: McKinsey State of AI 2025, Deloitte Connected Consumer 2025

2

실행 압박과 신뢰의 중요성

Gartner 조사에 따르면 77%의 서비스 및 지원 리더들이 AI 배포에 대한 경영진의 압박을 느끼고 있습니다.

출처: Gartner Customer Service Survey 2025

3

거버넌스가 성공 열쇠

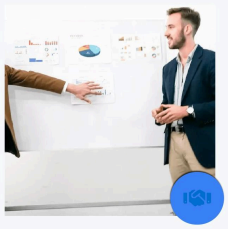
McKinsey 조사에 따르면 AI 거버넌스를 갖춘 기업이 74% 더 높은 성과를 보이며 투자 대비 수익을 실현합니다.

출처: McKinsey, 2024 AI 도입 성과 글로벌 설문조사

> 혁신적인 기술보다 더 중요한 것은 고객과의 신뢰 관계입니다
>
> NIST AI RMF, 2024

27

결론



신뢰

Deloitte(2024)에 따르면, 71%의 기업이 혁신과 신뢰 사이의 균형이 가장 큰 도전이라고 응답했습니다.

신뢰 중심의 AI 거버넌스가 경쟁 우위의 핵심입니다.

McKinsey(2024)에 따르면, 신뢰 기반 AI 전략을 구축한 기업은 고객 평생가치(CLV)가 평균 65% 증가했습니다.

출처: Deloitte TrustID™ Survey(2024), McKinsey Global Survey on AI(2024)

NIST AI RMF: "신뢰는 설계의 시작점이어야 합니다"

28

	Memo	

	Memo	

	Memo	

	Memo	

	Memo	

2025 공동 학술대회

Track 2



Track 2-1

AI 시대의 정보 보안과 거버넌스

최경호
(주)씨커스 대표



Track 2-2

토스뱅크 정보보호 거버넌스 사례

이정하
토스뱅크 CISO



Track 2-3

AI-to-AI 보안의 시대 - 차세대 디지털 신뢰 인프라 제안

이태훈
(주) 매드업 테크리드



Track 2-4

다양한 내부통제에도 불구하고 반복되는 IT 미비점의 구조적 원인 분석: 3개 금융회사를 IT내부감사 사례를 중심으로

김재영
딜로이트 안진회계법인 이사



Track 2-5

최신 보안 위협 대응을 위해 무엇을 할 것인가?

박형근
시큐리티플러스 커뮤니티 리더

	Memo	

Track 2-1

AI 시대의 정보 보안과 거버넌스



최경호
(주)씨커스 대표

발표개요

AI 시대의 보안 패러다임 변화와 거버넌스에 대응하기 위한 방법

이력

- 주식회사 씨커스 CTO
- 클라우드 보안 프로덕트 매니저(FRIIM CNAPP)
- 중대형 SI사업(공공, 금융) 다수 수행(PM, TA, DA, 개발PL, 보안 컨설팅 등)
- 공공 클라우드 이관 사업 다수 수행(프라이빗, 하이브리드 등)



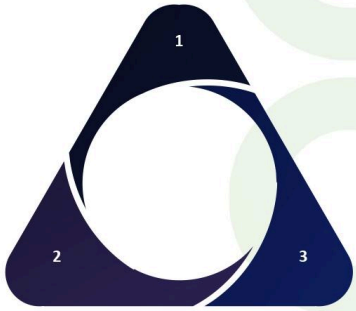
목차

1. AI 리스크 환경과 거버넌스 패러다임의 전환
2. AI 시스템의 근본적 취약점 및 감사 난제
3. ISO 42001 기반 통제 환경 구축
4. AI Sandbox와 실시간 감사
5. AI 환경별 보안 사고 사례 및 통제 전략
6. IT 감사인의 GRC AI 리더십

2

AI 발전에 따른 기존 거버넌스 활용 어려움

확률에 따른 예측 불가한 결과



시스템 자율성 증대

불확실한 신뢰성

3

COBIT (Control Objective for Information System and Related Technology)

Start with Strategic Alignment

Manage Your Risk

Measure Performance

Implement Security Protocols

Ensure Accountability



https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2025/cobit-a-practical-guide-for-ai-governance

4

COBIT 기반 AI 통제 목표 (1/2)

구분	카테고리	내용	비고
기존 COBIT 측정 및 통제 (정적)	애플리케이션 무결성	악의적 입력(프롬프트 인젝션)에도 모델이 의도된 안전 가이드라인을 벗어나지 않도록 행동의 일관성을 유지하는지 측정	
	접근 통제	AI가 자율적으로 목표를 추구할 때, 부여된 API나 시스템 Tool을 의도된 범위 내에서만 사용하도록 실시간 행동을 감시하고 제한	
	데이터 무결성	시간이 지남에 따라 모델의 성능이나 입력 데이터 분포가 변하여 모델의 행동이 악화되는 현상을 지속적으로 모니터링하고 경고	
윤리성 관련 통제	개인정보 보호	출력 결과에서 학습 데이터에 포함된 개인 정보나 민감한 속성이 유출되지 않도록 윤리적 위험을 직접 측정	
	결과의 정확성	인종, 성별 등 민감 속성에 대해 차별적이거나 편향된 출력을 생성하는지 자동화된 벤치마크를 통해 측정	
	컴플라이언스 규정 준수 확인	모델이 생성한 정보의 허위 여부를 실시간으로 측정하고 이로 인해 발생한 문제에 대한 책임을 추적할 수 있는 메커니즘 구축	

5

COBIT 기반 AI 통제 목표 (2/2)



Sandbox를 활용하여 거버넌스 강화 방안 제시

6

AI 시스템의 근본적 취약점 및 감사 난제 (1/2)

증폭 현상 (Amplification)

학습 데이터 내 포함된 편향 정보로 인해 극단적이고 차별적 결과 생성 가능

편향 전파 (Bias Inheritance)

초기 학습 및 파인 튜닝 과정에서 입력된 데이터셋에 포함된 경우 새로운 편향이 생길 수 있음

할루시네이션 (Hallucination)

정확하지 않은 결과 데이터를 설득력있게 제공하여 재무적, 법적 책임을 유발할 수 있음

7

AI 시스템의 근본적 취약점 및 감사 난제 (2/2)

진단 불가능 (Auditing)

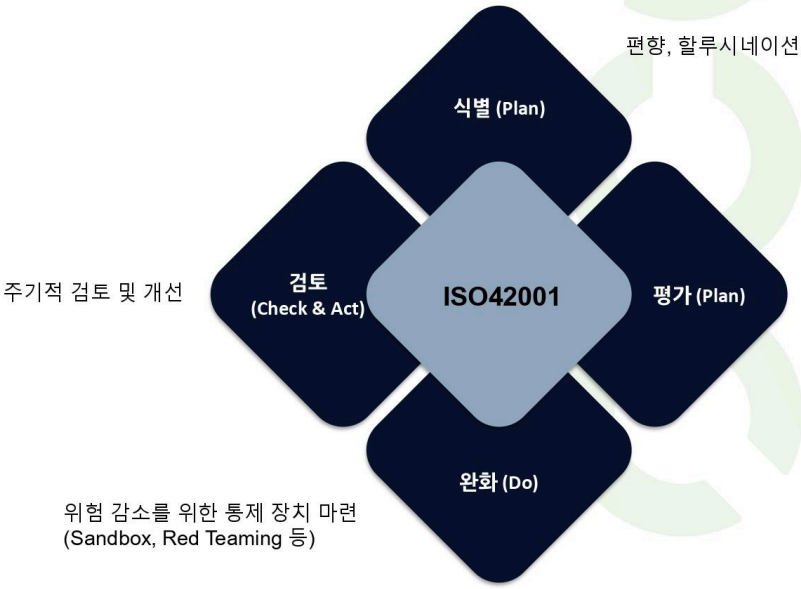
LLM의 맥락 의존성으로 인한 결정 과정의 불투명성, 블랙박스 모델에 대한 외부 감사 한계

AI 공급망 위험

복잡하게 역인 기술 스택 및 공급망에 의존하고있어 의부 데이터셋, API, Repository로 인한 문제 발생 가능

8

ISO 42001 기반 통제 환경 구축



ISO 42001 표준 준수는 기업이 AI 시스템의 윤리적/사회적 위험을 체계적으로 관리하고 있음을 외부에 입증할 수 있음 9

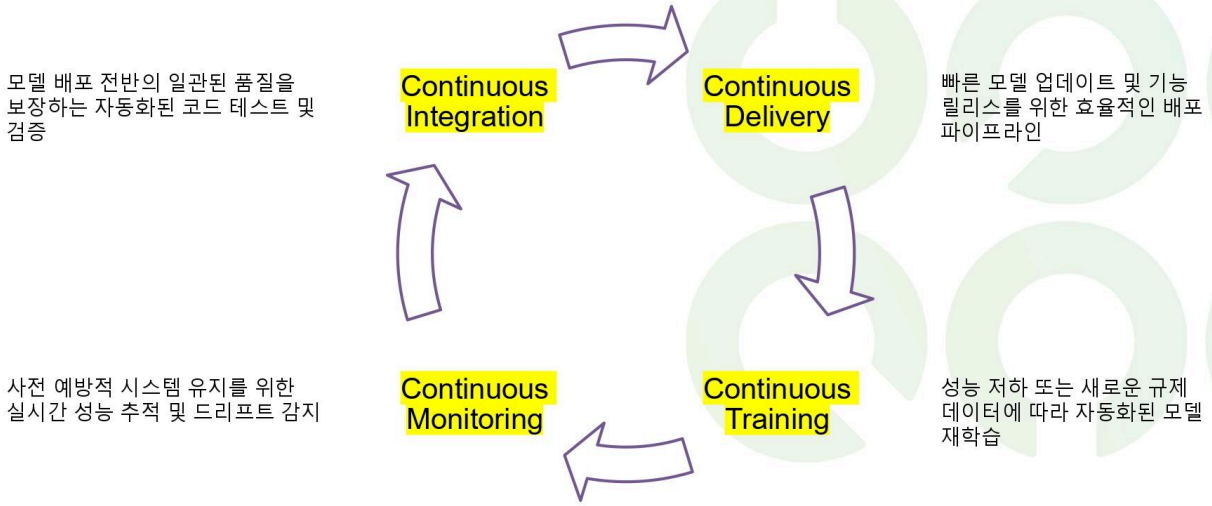
전 생애주기 Accountability 확보 방안

생애주기 단계	핵심 통제 목표	Accountability 확보 방안
개발 및 학습	데이터 윤리 및 편향 통제	• 데이터 Provenance (출처 추적) 기록 및 편향 영향 평가(AIIA) 의무화 • 사용한 모든 학습 데이터셋 및 전처리 과정의 버전 관리 및 기록 보존
배포 및 운영	실시간 행동 및 출력 통제	• LLM의 모든 프롬프트, API 호출, 응답에 대한 Audit Log (감사 기록) 보존 • 비정상적 행동 감지 시 즉시 차단 및 알림 책임 부여
모니터링	성능 및 신뢰성 유지	• Model Drift (모델 성능 저하) 및 Data Drift (데이터 분포 변화) 발생 시 재학습 • 폐기 결정에 대한 기준 및 책임 명확화
폐기	민감 정보 및 지식 소멸	• 모델 가중치, 학습 데이터, 생성된 로그 기록 등을 보안 정책에 따라 안전하게 파기했는지 확인하고 관련 기록을 보존

- C-Level : AI 위험 관리를 최고리스크책임자(CRO) 관할 영역으로 통합하는 등 거버넌스 구조 조정
- AI Owner : 모델의 편향성 검증, 데이터 사용 승인, 운영 통제 등 핵심 의사 결정에 대한 최종 책임

10

AI 모델 관리를 위한 효과적인 프로세스



11

AI Sandbox (Pre-Deployment)

개요	모델이 배포되기 전에, 학습 데이터를 통한 잠재적 위험을 차단하고 모델의 건전성을 확보하는 데 중점
목표	• 데이터 중독(Data Poisoning) 방어 및 편향 측정 • 악의적인 공격자가 학습 데이터에 오염된 샘플을 삽입하여 모델의 신뢰성을 손상시키는 것을 차단
구현 방법	1. Differential Privacy (DP) 환경 구축 민감한 개인 정보(PII)가 포함된 데이터셋을 학습에 사용할 때, 개별 데이터 샘플의 영향을 최소화하는 통계적 노이즈 주입 이를 통해 모델의 출력을 분석하더라도 특정 개인이 학습 데이터에 포함되었는지 여부를 추론하기 어렵게 만들어 개인 정보 유출 위험 차단 2. 외부 데이터셋 무결성 검증 외부에서 가져온 데이터셋에 대해 악성 코드나 고의적인 오염 흔적이 없는지 자동화된 도구를 이용해 철저히 검증 3. 편향 측정 다양한 민감 속성(성별, 인종 등)에 대해 모델이 편향된 결과를 내는지 학습 전후에 시뮬레이션을 통해 측정 및 보고

12

AI Sandbox (Runtime Protection)

개요	운영 단계의 AI Sandbox는 사용자가 LLM을 실제로 사용할 때 발생하는 위험, 특히 내부 정보 유출과 악성 입력/출력을 실시간으로 감시하고 차단하는 필터링 게이트웨이 역할 수행
목표	• 내부 데이터 유출 방지 및 연동 통제 • 불필요/악성 정보 반입 차단 및 유효성 검사 • 동적 편향 감사 방법론 (Red Teaming의 확장)
구현 방법	1. 입력(Input) 필터링 2. 출력(Output) 마스킹 3. Input/Output 유효성 검사 핵심 4. 자동화된 테스트

13

AI Sandbox 핵심 장점

민감 정보 유출 원천 차단 (Confidentiality)

모델의 무결성 및 강건성 확보 (Integrity & Robustness)

위험 행동의 자율적 통제 및 격리 (Containment)

책임 추적 및 감사 가능성 확보 (Accountability & Auditability)

14

AI 환경별 보안 사고 사례 (1/3)

환경 : RAG 및 일반 LLM 배포 환경 (Data Leakage & Disclosure)

사례 1 : 제조사 기밀 정보 ChatGPT 유출

- 개요 : 직원들이 통제되지 않은 외부 LLM 서비스(ChatGPT)에 내부 소스 코드, 회의록 등 기업 기밀 정보를 입력하여 요약이나 검토 요청
- 보안 문제 : 입력된 데이터가 외부 서비스의 서버로 전송되어 저장되거나 모델 학습에 활용될 가능성이 생기면서 내부 정보 유출 발생. 이는 OWASP의 Sensitive Information Disclosure (민감 정보 노출) 취약점에 해당

15

AI 환경별 보안 사고 사례 (2/3)

환경 : RAG 및 일반 LLM 배포 환경 (Data Leakage & Disclosure)

사례 2: OpenAI 결제 정보 유출 버그

- 개요 : OpenAI의 Redis 서버 버그로 인해 일부 유료 사용자의 채팅 기록 제목과 결제 정보(신용카드 번호 일부, 유효기간)가 다른 사용자에게 노출
- 보안 문제 : LLM 운영 인프라 자체의 보안 취약점이나 데이터 격리 실패로 인해 민감한 개인 정보(PII)가 유출 / LLM 모델 자체의 문제가 아닌, 모델을 둘러싼 인프라 및 서비스 관리 문

16

AI 환경별 보안 사고 사례 (3/3)

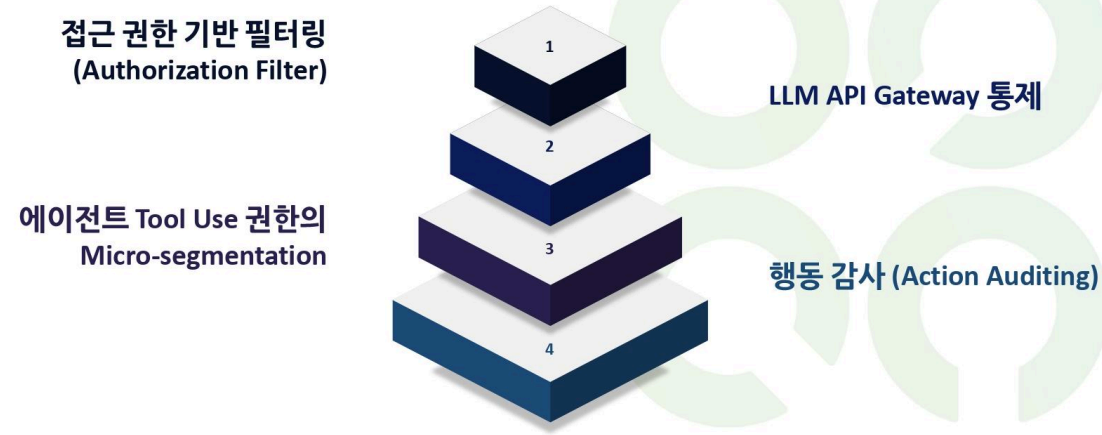
환경 : Agentic AI 환경 (Excessive Agency & Malicious Action)

사례 2: 에어 캐나다 챗봇 허위 정보 제공

- 개요 : 에어 캐나다의 챗봇이 고객에게 잘못된 할인 정책을 안내했고, 이는 고객이 회사를 상대로 소송을 제기하게 만드는 법적 책임 이어짐
- 보안 문제 : LLM이 부여된 과도한 대행(Excessive Agency)과 할루시네이션이 결합되어, 모델의 오류가 단순한 정보 오류를 넘어 실제 비즈니스 결과(재무적, 법적 손실)를 초래

17

AI 보안 사고 통제 전략



18

IT 감사인의 AI GRC 리더십

구분	내용
통제 대상의 변화	통제 대상이 정적인 인프라(서버, 네트워크)에서 확률적이고 자율적인 모델의 행동(Behavior)으로 이동
위험의 형태	위험이 단순히 해킹이나 데이터 유실을 넘어, 모델의 편향성, 윤리 위반, 할루시네이션 등 신뢰와 책임에 관한 문제로 확장
새로운 요구 사항	단순히 외부 공격을 막는 것을 넘어, 모델이 내부에서 해로운 행동을 하지 않도록 통제하고 윤리적 원칙을 준수하는지를 보장하는 것이 핵심
필요성	• IT 감사인은 기존 COBIT이나 ISO 기반의 체계적인 통제 프레임워크 설계에 대한 전문성을 갖추고 있음 • 이 전문성을 활용하여 ISO 42001과 같은 AI 거버넌스 표준을 기업 전체에 효율적으로 적용 • AI 위험을 ERM에 통합하는 역할 주도
주도해야 할 기술	• AI Sandbox (학습/운영 단계의 격리 및 필터링) • Dynamic Auditing (Red Teaming, Prompt Engineering 편향/취약점 검증) • 기타 새로운 동적 통제 기술

19

	Memo	

Track 2-2

토스뱅크 정보보호 거버넌스 사례



이정하
토스뱅크 CISO

발표개요

토스뱅크는 고객의 금융 경험 혁신을 최우선 가치로 삼고, 안정적고 효율적인 정보보호 거버넌스 체계를 구축하여 비즈니스 목표를 달성하고 있습니다.

최근 다양한 산업분야에서 보안사고가 발생하고 있는 상황에서 토스뱅크의 정보보호 활동을 소개함으로써 정보보호 업무에 도움이 되었으면 합니다.

이력

- 토스뱅크 정보보호 디비전 본부장
- 정보보호시스템감사통제협회 보안부문 부회장
- 사이버보안전문단

	Memo	

Track 2-3

AI-to-AI 보안의 시대 - 차세대 디지털 신뢰 인프라 제안



이태훈
(주)매드업 테크리드

발표개요

1. 오프닝
2. 보안 패러다임 전환
3. 기존 통제의 공백
4. 제안: Trust Layer 한장
5. Trust Receipt (감사 영수증)
6. 아키텍처 (컴플라이언스 버전)
7. 산업별 케이스 스터디
8. 거버넌스 운영모델
9. 바로 쓸수 있는 내부 감사 체크리스트 예시
10. 클로징

이력

- (주) 매드업 테크리드
- 前 (주) 세인트블록 책임연구원
- 前 Cyphersys Pty Ltd - Chief Producer (호주근무)
- 前 (주) 현대해상화재보험 근무
- 前 부산 핀테크허브 - 강사 역임
- 前 패스트캠퍼스 - 강사 역임



발표자 소개

이태훈 (CISA, Since 2022)

來 엠지알브이, Head of Engineering
現 매드업, Techlead
前 Cyphersys Pty Ltd - Australia
前 현대해상

부산 핀테크 허브, 패스트캠퍼스 등 출강

2

발표에 사용되는 이미지, 폰트, 템플릿

폰트 – Arial
템플릿 – ISACA Korea 제공
이미지 – Gemini Pro 2.5 생성

3

Castle & Moat



4

Table of Contents

1. Philosophy
2. 문제정의 (Why Now)
3. 제안: A2A Trust Broker
4. 코어구조: 6-Layer 신뢰 인프라
5. 거버넌스 & 감사 관점 정리
6. 확장 시나리오(AI 신뢰중개가 커졌을 때)
7. Closing

5

Philosophy

1. AI의 흔적(Transactions) 들도 결국 기록되어야 한다.
2. 보안보다 선행하는 건 신뢰구조.
3. 표준이 없으면 감사가 없다.

6

AI의 흔적(Transactions) 들도 결국 기록되어야 한다.

→ **사람끼리 송금하면? 카드를 사용하면?**
지울수 없는 기록들이 많은 곳에 새겨짐

→ **"이 고객 정보 요약좀 부탁해~"**
현재 그 어디에도 기록이 남지 않음

→ **나중에, 이 정보가 어디까지 노출되어 있지?**
아무도 모름

7

보안보다 선행하는 건 신뢰구조

→ **우리가 CISA/CISSP에서 학습한 보안 모델은 이제 AI 에이전트까지로 확장되어야 합니다.**
누구를 믿을 것인가.. 그것이 사람인가요? 에이전트 인가요?

→ **암호화, 인증, 방화벽도 중요하다.**
누구에게 / 무엇을 / 어디까지 열지 신뢰 구조가 먼저 정의되어야 보안이 의미를 갖습니다.

→ **요약하자면,**
보안 기술을 강화하기 전에 '**누가 누구를 믿고 무엇을 공유하는가**'의 지도를 먼저 그려야 합니다

8

표준이 없으면 감사가 없다

→ **거대한 표준을 만드는 것은 상당한 시간과 노력이 들어간다.**
조직 내부에서라도 '우리 포맷'이라는 작은 프레임워크를 먼저 가져야 한다.

→ **포맷이 다르면 감사가 어려워진다.**
조직용 미니 표준(Who / Why / What / When) 을 먼저 찍자.

9

1. Philosophy
2. **문제정의 (Why Now)**
3. 제안: A2A Trust Broker
4. 코어구조: 6-Layer 신뢰 인프라
5. 거버넌스 & 감사 관점 정리
6. AI 신뢰중개가 커졌을 때
7. Closing

2.1.
사람→시스템 시대에서 AI→AI 시대로의 전환

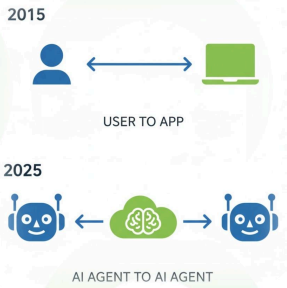
2.2
주체가 사람이 아닐 때 생기는 5가지 공백

2.3
감사 기준을 빠르게 정해야 하는 이유

10

2.1 사람→시스템 시대에서 AI→AI 시대로의 전환

이제 사용자가 아니라 에이전트가 호출합니다.



- 과거: 사람 → 앱 (API)
- 지금/미래: AI 에이전트 → 다른 AI/시스템
- 문제: 주체가 사람이 아니면 기존 접근통제를 생각해 보았나..?

11

그래서, 주체가 사람이 아닌 것이 뭐가 문제?

- 권한의 상속폭이 너무 크다.
  - 최소권한 정책을 맞추기 쉽지 않음
- 의도가 남지 않는다.
  - 사람은 "고객의 목록을 뽑으라" 시켰지만,
  - 에이전트는 추가적인 API를 호출할 수 있음. 각 호출의 의도(intent)가 남지 않음
- 파생호출이 보안 레이어를 우회함
  - A2A 첫 호출 때 WAF/IPS/접근통제를 다 거쳐 감.
  - 에이전트다 내부서비스끼리 부를 때 추가 레이어가 필요할 수 있음
- 행동의 일관성
  - 사람의 행동은 일관적이다.
  - 에이전트는 모델과 프롬프트 버전이 바뀌면 같은 권한으로 다른 행동을 할 수 있다.
  - 이때 아무것도 남지 않는다면...?

12

2.2 주체가 사람이 아닐 때 생기는 5가지 공백

신원 공백
- 호출 주체: 분명 사람의 토큰을 가지고 입장했다.
- 실제 실행자: 실제로는 Agent V3가 호출함

목적 공백
- 최초 요청 목적은 있음 (예: “고객 세그먼트 뽑기”)
- 에이전트가 파생해서 부른 호출의 **목적 라벨이 없음**
- **의도기반 감사가 어려움**

범위 제한 공백
- 사람 권한을 **그대로 가져가서**
- 실제 작업보다 **더 넓은 데이터**를 들여다봄

책임 공백
- 사고 나면 누구 책임이지.
- 사용자 vs 에이전트 vs 에이전트 만든 팀?
- 지금도 모두를 전부 감사하지만, 지금 같은 정책 과도기에는 감사 자체가 너무 어렵다.

가시성 공백
- 1차 호출은 보안/로그를 타는데
- 2·3차 내부 호출은 **가시성이 떨어짐**

13

2.3 감사 기준을 빠르게 정해야 하는 이유

로그
시간이 지남에 따라 그 가치는 점점 떨어지지만, 감사를 위해 반드시 필요한 정보이자, 사고 발생시 역추적을 위한 중요한 단서.

포인트
- 각 서비스가 **제각각** 로그는 남기고 있음
- 포맷이라도 통일 시켜야 함.
- 나중에 사고 났을 때 **어떤 호출이 문제였는지 역추적이 안 됨**

사내 표준
표준이 없으면 감사가 개발 프로젝트가 됩니다. 감사가 매달 ‘스크립트 짜는 일’이 되면 이미 진 겁니다.

14

1. Philosophy
2. 문제정의 (Why Now)
3. **제안: A2A Trust Broker**
4. 코어구조: 6-Layer 신뢰 인프라
5. 거버넌스 & 감사 관점 정리
6. AI 신뢰중개가 커졌을 때
7. Closing

3.1
선(先) 브로커, 후(後) 호출

3.2
브로커가 대신 해주는 6가지 일

3.3
AI 에이전트 중개 플랫폼으로 확장되는 그림

15

3.1 선(先) 브로커, 후(後) 호출

에이전트 사이에 아주 얇은 브로커를 둔다.

지금: AI → (바로) 다른 AI/API
제안: **AI → Trust Broker → AI/API**
이유: 브로커를 거쳐가야 **정책 입히고, 기록 하고, 필터 가능**

💡 에이전트가 10개로 늘어나도 **브로커 규칙만** 바꾸면 된다.

16

3.2 브로커가 대신해주는 6가지 일

신원 확인
- 에이전트 전용 신원 확인 레이어

라우팅 및 버전 중재
- 이 호출은 v3 버전을 호출.
- 이 호출은 특정 LLM 모델을 호출 등

정책 및 동의 적용
- 특정 목적/범위만 허용하는 등

감사영수증 발행
- 감사로그 작성
- Who / Why / What / When / Proof 남김

데이터 방화벽
- PII / 민감정보 들어오면 마스킹/비식별

비용 및 리스크 제한
- 많이 / 자주 호출하면 컷 할 수 있어야 함

17

3.3 AI 에이전트 중개 플랫폼으로 확장되는 그림

오늘은 보안, 내일은 에이전트 마켓

- 1단계: 내부 AI들만 브로커 경유
- 2단계: 외부/파트너 에이전트도 브로커 통해 접속
- 3단계: **신뢰된 에이전트**만 노출하는 중개/마켓플레이스로 확장
- 부가서비스: 인증, 성능지표, 호출분석, 유료 프리미엄 라우팅

💡 결국 '누구를 믿고, 어떤 에이전트를 써도 되는가'를 결정하는 **중앙 신뢰기관 역할**까지 갈 수 있다.

18

1. Philosophy
2. 문제정의 (Why Now)
3. 제안: A2A Trust Broker
4. **코어구조: 6-Layer 신뢰 인프라**
5. 거버넌스 & 감사 관점 정리
6. AI 신뢰중개가 커졌을 때
7. Closing

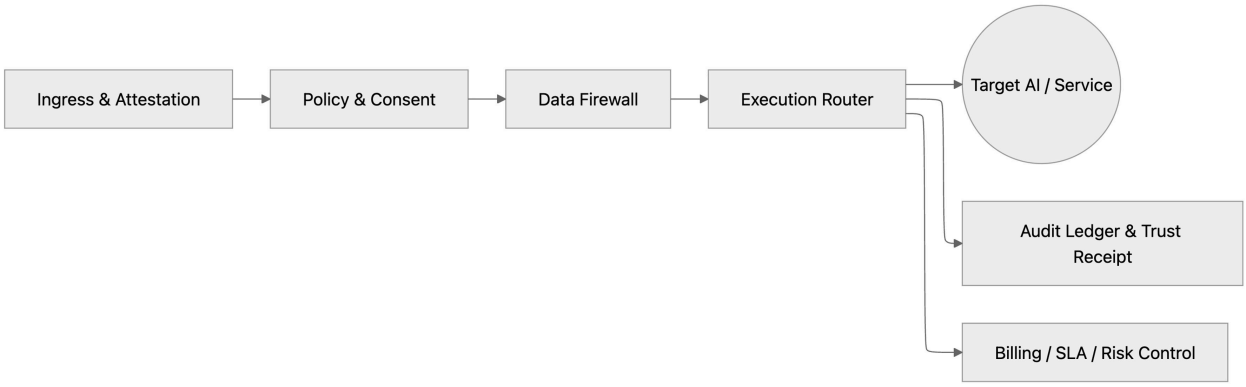
4.1
6 Layers

4.2
들어오기 전 방어하는 3개 레이어

4.3
실행, 기록, 제한을 위한 3개 레이어

19

4.1 6 Layers



- 누군지 확인 → 허용되는 호출인지 보고 → 데이터를 줄이고 → 어디로 보낼지 정하고 → 증빙을 남기고 → 비용/리스크를 제한
- 이 순서가 뒤집히면 **사후 포렌식(After Control)**

20

4.2 들어오기 전 방어하는 3개 레이어

Ingress & Attestation (누가 들어왔는가)
Policy & Consent Orchestrator (적법한 호출인가)
Data Firewall (과다 노출을 먼저 막는다)

- 에이전트/모델/툴의 신원·버전을 받아둔다
- 에이전트 전용 토큰을 사용하거나, 에이전트 신분을 받아야 한다.

- 목적/업무/데이터등급 등으로 허용/거절/조건부
- 사람한테는 OK이지만 에이전트 호출은 No 일수 있음
- 내부정책(예: 고객 P0는 외부모델 사용금지) 등을 적용가능

- PII/민감 토큰 탐지 → 마스킹/비식별/샘플링
- 프롬프트/컨텍스트에 섞여 들어온 고객정보를 실행 전에 필터
- 이 레이어는 현재 매우 필요한 상태

2

4.3 실행/기록/제한 구간

Execution Router (어떤 에이전트/버전으로 보낼 것인가)
Audit Ledger & Trust Receipt (기록을 증빙으로 만든다)
Billing / SLA / Risk Control (비용·속도·리스크 상한)

- 모델, 버전, 내외부를 선정
- 성능/비용/국가·데이터 규제 기준으로 라우팅 룰을 중앙에서 가능
- 모델이 바뀌어도 이 레이어만 있다면, 예측밖의 행동을 하지 않는다.

- 기준에 맞추어 로그 및 증빙기록을 남김
- 로그를 넘어 감사 영수증, 블록체인에 기록하면 위변조 불가

- 지나치게 많은 호출, 비정상 속도 제한
- 호출량/비용/리스크를 한 레이어에서 제한하고. 한 화면에서 모니터링 가능

22

1. Philosophy
2. 문제정의 (Why Now)
3. 제안: A2A Trust Broker
4. 코어구조: 6-Layer 신뢰 인프라
5. **거버넌스 & 감사 관점 정리**
6. AI 신뢰중개가 커졌을 때
7. Closing

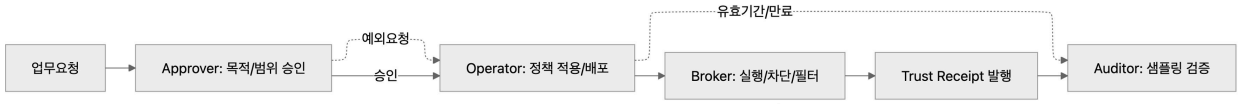
5.1
RACI: 누가 소유/승인/운영/감사하는가

5.2
정책 4종 세트

5.3
샘플링 감사 넣는 법

23

5.1 RACI: 누가 소유/승인/운영/감사하는가



역할이 없으면 통제도 없다.

Owner(소유): 데이터 등급 지정, 정책 소유, 보존/파기 승인
Approver(승인): 호출 목적/범위 승인(업무책임자), 예외 만료 관리
Operator(운영): 브로커/정책 배포, 차단 및 해제, 모니터링
Auditor(감사): Trust Receipt 샘플링/증빙 검증, 재발 방지 조치

Responsible
AI 호출의 **주인**은 호출자가 아니라 **데이터 Owner**

24

5.2 정책 4종 세트 (목적, 등급, TTL, 재전달금지)

Purpose(목적 제한)
업무 목적 태그 필수. 목적 외 사용 자동 차단.

Data Class(데이터 등급)
P0(민감)·P1(사내)·P2(공개) + 허용 매트릭스.

TTL(체류시간)
호출 맥락/캐시/로그의 보존시간 및 폐기 규칙.

Downstream Ban(재사용 및 재전달 금지)
2·3차 호출 시 범위·출처 태그 유지.

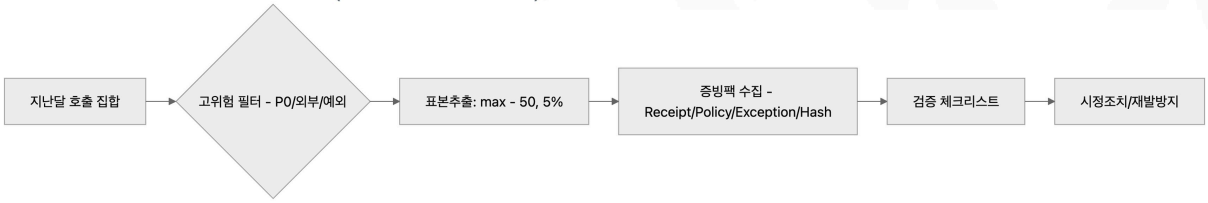
25

5.3 AI 호출도 샘플링 감사 대상

감사는 프로젝트가 아니라 루틴입니다.

예시

- 대상 지정: 월 1회, 고위험 호출(P0, 외부전송, 예외 건) 우선
- 표본추출: max(50건, 지난달 고위험 호출의 5%) (상한 500건)
- 증빙팩: 각 표본의 Trust Receipt + 정책버전 + 예외서명 + 로그해시
- 검증항목: 목적 일치, 등급·TTL 준수, downstream 금지 준수, 우회 호출 0건
- 시정조치: 위반유형 분류(정책/운영/교육), 만료·정책 보강, 재발 방지 기한



26

1. Philosophy
2. 문제정의 (Why Now)
3. 제안: A2A Trust Broker
4. 코어구조: 6-Layer 신뢰 인프라
5. 거버넌스 & 감사 관점 정리
6. **AI 신뢰중개가 커졌을 때**
7. Closing

6.1
개방형 생태계로 열릴 때: 오픈소스/표준 포인트

6.2
반드시 올 미래 (연결이 돈이 되는 순간)

27

6.1개방형 생태계로 열릴 때: 오픈소스/표준 포인트

왜 개방형 생태계로 열릴까?

기업마다 에이전트가 다 다르면 감사가 불가능하니까, 결국 요청·응답·영수증 포맷은 열려야 함. MCP와 같다.

감사포인트 #1

이 조직이 공개스키마를 따르나?

감사포인트 #2

내부 정책이 외부 OSS 정책 엔진(가이드라인) 을 따르고 있나?

감사포인트 #3

오픈소스 브로커를 도입했어도 영수증 저장위치 보존 주체는 여전히 우리쪽인가?

28

메세지

오픈소스여도 감사는 닫혀 있어야 한다.

29

6.2 반드시 올 미래 (연결이 돈이 되는 순간)

현재개념
내부 에이전트가 외부 에이전트를 호출한다

미래개념
외부 에이전트가 우리 데이터를 비용을 지불하고 이전 받아 사용한다.

과금, 정산, 책임이 얽히게 된다.

30

6.2 반드시 올 미래 (연결이 돈이 되는 순간)

감사 포인트 #1
요청자와 결제자 일치 여부 (Shadow AI 사용 방지)

감사 포인트 #2
데이터 등급별 단가/정책 차별이 규정과 합치하는지

감사 포인트 #3
마켓플레이스가 발행한 영수증과 내부 브로커 영수증이 일치하는지 (이중장부 탐지)

감사 포인트 #4
재판매/다운스트림 호출이 원 정책을 계승했는지 ('목적 외 사용 금지' 원칙 준수)

31

1. Philosophy
2. 문제정의 (Why Now)
3. 제안: A2A Trust Broker
4. 코어구조: 6-Layer 신뢰 인프라
5. 거버넌스 & 감사 관점 정리
6. AI 신뢰중개가 커졌을 때
7. **Closing**

32

7. Closing

- 기록되지 않은 AI는 통제할 수 없다
- 신뢰를 먼저 깔아야 보안을 얹을 수 있다
- 오늘 잡은 이 최소 표준이 내년 감사의 기준이 될 것이다

33

7. Closing

- 기록되지 않은 AI는 통제할 수 없다
- 신뢰를 먼저 깔아야 보안을 얹을 수 있다
- 최소 표준이 내년 감사를 편안하게 만들 것 이다.

34

Track 2-4

다양한 내부통제에도 불구하고 반복되는 IT 미비점의 구조적 원인 분석: 3개 금융회사를 IT내부감사 사례를 중심으로



김재영
딜로이트 안진회계법인 이사

발표개요

지속적으로 IT compliance 요건이 강화되고, 금융회사를 대상으로는 책무구조도 등 다양한 요구사항이 지속적으로 발생하고 있습니다. 본 발표는 2025년 수행된3개 금융회사에 대한 IT 내부감사 결과를 비교·분석하여, 근본적 원인을 규명하고 실무적 우선순위 권고를 제시합니다.

이력

- 딜로이트 안진회계법인 이사
- ISACA AI Audit 연구회 연구위원

다양한 내부통제에도 불구하고 반복되는 IT 미비점의 구조적 원인 분석
: 3개 금융회사의 IT내부감사 사례를 중심으로

김재영
딜로이트 안진회계법인 | 이사

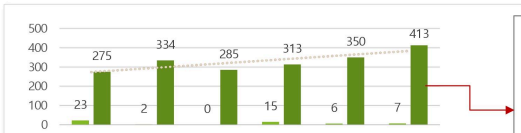
Confidential. For internal use only.

ISACA Korea Chapter

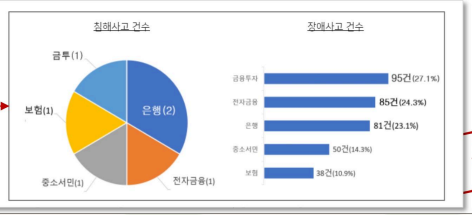
1.1. 배경 - 외부 환경 변화

전자금융 사고증가

IT침해사고*의 경우 '14년부터 전체 금융권의 보안대책을 단계적으로 강화하여 대형 침해사고로 인한 소비자 피해 사례는 발생하지 않음
IT장애사고**의 경우 비대면 거래 증가 및 증권시장 활황 등에 따른 이용자 폭증으로 인한 서비스 지연 사례가 있었고, 차세대시스템 구축 및 오픈 뱅킹 등의 신규서비스 출시 과정에서 프로그램 오류 적용 등으로 인해 일부 소비자 피해가 발생한 사례 有



구분	2017년	2018년	2019년	2020년	2021년	2022년
침해사고	23건	2건	0건	15건	6건	7건
장애사고	275건	334건	285건	313건	350건	413



금융권역별 전자금융사고 발생현황

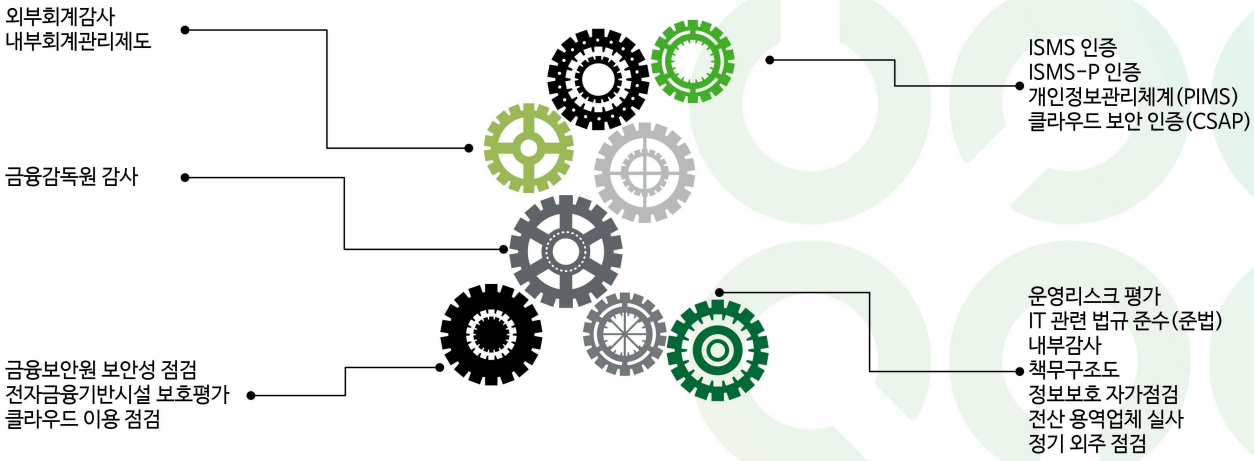
금융IT 안전성 강화 가이드라인 (23.11)	
전산시스템 성능관리	• 성능관리 임계치 설정 및 대응전략 수립 • 대형이벤트 유입량 분석 및 예측 • 성능관리 비상대책 마련
비상대책 수립·운용	• 비상훈련 실효성 강화 및 환류체계 • 핵심업무 선정 및 부서별 역할 명확화 • 업무지속성 확보 점검 및 시스템 구축
프로그램 통제	• 제3자 검증 통제기능강화 • 테스트 역량 강화(전담화, 자동화) • 프로그램 통제 관리 및 점검 강화

금감원 IT 감사 가이드라인 (25.2)	
3단계 IT내부통제 체계 구축	• 1단계(IT현업) : IT내부통제 방안수립 및 이행 • 2단계(IT자체감사) : IT조직 內 IT내부통제 자체 점검 • 3단계(IT감사) : 감사조직의 IT감사
IT내부통제 범위 및 주체 명확화	• IT/정보보호/디지털(AI) 포함한 IT업무 전반 • IT영역별 최고책임자가 소관 IT업무에 대한 내부통제 활동 수행
표준 IT감사 방법론 제시	• IT리스크 평가 기반 IT감사절차(계획/실시/보고) • IT감사인의 전문성 및 독립성 확보 (필요시 외부위탁 가능)

금감원 2025년 디지털·IT부문 검사 방향 (25.2)	
중요 테마별 중점 검사사항	• 데이터 처리 실태 • IT 운영 및 보안통제 등
중층적 IT내부통제 및 사전예방적 검사 강화	• 금융회사가 스스로 IT 리스크를 평가 • 3단계 IT내부통제 수행 • IT 자율시정 확대
IT신기술 관련 리스크 관리 강화 및 운영 복원력 제고	• 제3자 IT리스크, 보안통제 사각지대 • 회복탄력성 검증, 재해복구센터 개선, IT사고 대응체계 개선

2

1.2. 배경 - 금융회사의 IT 관련 내외부 통제 절차

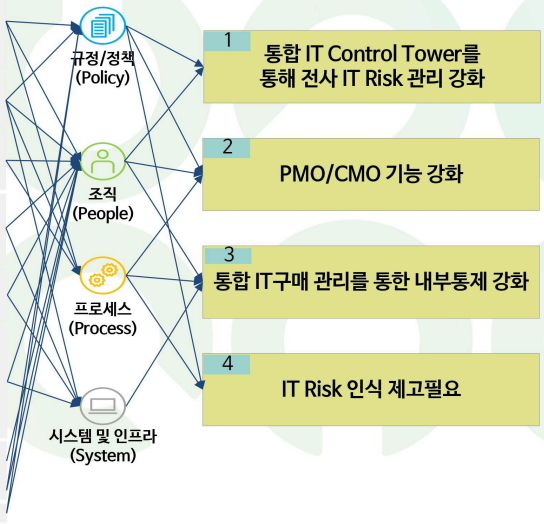


많은 내부통제 점검을 수행함에도 사고가 지속적으로 발생하는 이유는?

3

2.1. IT 내부감사 사례 – A 금융회사

Key Finding	시사점
계약 변경 시 변경의 적정성 검토 및 이력관리 필요	IT투자 및비용지출 통제강화 주기적비용최적화활동필요
투자 효율성을 위한 예정가격 산정 및 가격평가 규정지침 제고	
IT 자산관리시스템 상 완전성 있는 자산관리 필요	
예정가격 검증 강화 필요	
외주업체 유지보수 검수확인서와 실제 수행이력 불일치	
납품대상 S/W, H/W 에 대한 검수확인서/납품확인서 완전성 미흡	
외주 정보 등록 현행화 필요	IT운영관련품질제고필요
오픈소스 SW의 라이선스 점검 및 취약점 점검절차 개선 필요	
비상상황 대응 매뉴얼 및 재해복구 센터 관리 미흡	
프로그램 변경 시 부적절한 테스트 증빙 첨부	전산일반통제 운영강화및 ITRisk관련인식제고필요
프로그램 변경 시 요청 프로세스 미흡	
서버라우팅처리기준미흡	
업무계/정보계 운영서버 내 파일 무작위 업로드 및 다운로드	전산응용통제취약점식별및 장애/오류예방적조치강화
운영상 불필요한 IT비상 화면 존재	
하드코딩을 통한 특정 사번 및 부서코드 예외처리	
IP 접근 통제 및 단말기 보안 미흡	정기활동대상외 외부침입취약점진단확대
외부 그룹웨어 평문 전송으로 ID/PW 노출	
과도한 내부 네트워크 정보 노출	
통합/대시보드 상면 권한관리 기준 수립 필요	내부정보과다노출건에대한 진단을통해 취약점개선
프로젝트 산출물을 통한 시스템 관리자 ID/PW 노출	
주요 시스템 보안 설정 강화 필요	
IP접근 통제 및 단말기 보안 미흡	
화면보호기 차단해제 프로세스 강화 필요	
실효성 있는 보안 모니터링 및 관리를 위한 프로세스 수립 필요	
내부감사 발견사항에 대한 부적절한 개선사항	Compliance통제강화
WTS, MTS 필수동의 절차 상 광고성 문구 제고 필요	거버넌스개선
AI관련 거버넌스 강화 필요	거버넌스변화관리강화
외부 규정 개정사항 미반영에 따른 내부 규정 현행화 필요	



4

2.2. IT 내부감사 사례 – B 금융회사

핵심영역	Key Finding	시사점
IT거버넌스 및 감사체계	IT 감사 범위·대상·절차 불명확 및 제한	전사 IT 거버넌스 및 감사체계 강화
	감사 인력·기간·독립성 부족	
	정보보호위원회 보고·의사결정 절차 미흡	
	전사 규정·지침 검토 및 사규화 절차 부재	
	IT사업 추진·구매 관련 기준 및 절차 부재	
보안접근통제 관리	방화벽·보안장비 설정 및 보안패치 미흡	정보보안 및 접근권한 관리 강화
	비밀번호 정책·설정 취약	
	접근권한·계정관리 미흡	
	외부 서비스·웹 보안 설정 미비	
	물리적 보안통제 취약	
개인정보 규제준수	개인정보 처리방침·동의서 불명확·누락·불일치	개인정보 처리 및 보호조치 강화
	개인정보파기절차 및 이행미흡	
	개인정보 안전성 확보조치 미흡	
	전자금융거래 법규 및 절차 부재	전자금융거래 및 법규준수 강화
	금융거래 제한대상자 거래방지 및 위반자 처리 절차 부재	
변경형상 프로젝트 관리	변경배포형상관리절차미흡	변경배포형상관리 절차강화
	프로젝트 종료 후 인수인계 및 문서화 미흡	
	오픈소스·SW 사용현황 관리 부재	
클라우드·외부서비스 관리	클라우드 서비스 업무 중요도·건전성 평가 미수행	클라우드·외부서비스 관리 기준 수립
	클라우드 계정관리·접근제어 취약	
	클라우드 로그·Config 이력 관리 부재	
	외주·전자금융보조업자 관리·평가 미수행	
재해복구·비상대응	비상대응계획 수립·운용 미흡	비상계획의 실효성 강화
	DR센터 이격·가용성 기준 미충족	
	핵심/비핵심 업무 미선정·절차 미수행	
업무분장·역할책임 구분	업무분장·역할책임 정의 미흡	조직 역할·책임 및 성과관리 수립
	인력 확보 방안 부재	
	개발/운영 업무분장 불명확	
장애성능 관리	대형 이벤트 부하테스트 미수행	장애성능관리 실효성 강화
	자원 임계치 보고 절차 부재	

To-Be 개선과제 도출

1. 자체감사 활동 강화
2. 내/외부 감사 대응 일원화
3. 팀별 Control Tower를 통한 IT Risk 관리 강화
4. PMO/CMO 기능 강화
5. IT 전략에 따른 조직 체계 마련
6. 인식 제고

5

2.3. IT 내부감사 사례 – C 금융회사

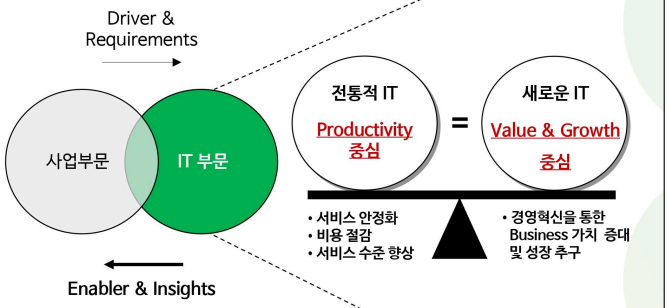
Key Finding	시사점
IT 자체감사 시 감사대상 범위의 완전성 부재	IT감사 체계강화
IT 자체감사인 평가에 대한 독립성 부재	개인정보 관리체계강화
개인정보 처리 관련 사규의 Compliance 요건 미충족	권한관리 체계강화
정보처리시스템의 권한 적정성 검토 절차 미비	데이터 변경절차강화
클라우드 관련 내규 부재	
데이터 변경에 대한 제3자 검증 미수행	정보보호 체계강화
일괄 작업 처리 시 공동 계정 사용 및 사후 검증 절차 부재	
외부 메일로의 반출 절차 확인 불가	
출력에 대한 로그가 적재되지 않음	
출력물에 대한 물리적 통제 미수행	
SLA 내역에 대한 Penalty 및 Incentive 정책의 유효성 부족	제3자 리스크 관리강화
도박/음란 사이트가 일부에 한정되어 차단됨	
프로그램 배포 시 안전성 조치(취약점 검사 등) 부재	프로그램 변경절차강화
오픈소스 점검 시 라이선스 및 취약점 검토 절차 미흡	
프로그램 등록, 변경, 폐기에 대한 제3자 검증 절차 부재	

To-Be 개선과제 도출

1. 전사 Control Tower 마련 필요
2. IT 감사 체계 강화
3. IT 위험에 대한 인식 부족
4. PMO/CMO 기능 강화

6

3.1. 주요 개선방안 - IT Risk 인식 제고필요



IT의 발전 방향

IT기반의 수익 창출 및 경쟁력 강화 기회 모색

- 현업의 IT 니즈에 대한 명확한 이해
- IT를 통한 경쟁 우위 확보 가능 요소에 대한 현업 리딩
- 현업의 추진안에 대한 IT 구현 가능성 검토 및 대안 제시

IT Risk에 대한 철저한 관리로 비즈니스 연속성 보장

- IT Risk전반에 대한 통제 (IT Risk관리는 민첩성을 떨어뜨리는 것이 아니라, 시행착오를 줄임으로서 민첩성을 제고하고, 비용 낭비를 최소화하는 목적)
- 최적의 기술적 대안 제시 및 정보자산 보호 방안 제시
- 현업의 IT 수요에 대한 올바른 의사결정 지원
- 현업의 IT수요와 전사 전략과의 연계성 평가
- 환경 변화에 능동적으로 대응할 수 있는 IT 조직 구성
- 연계 확보/강화를 위한 IT부문의 프로세스 명확화

7

3.2. 주요 개선방안 - 전사 IT Control Tower 마련

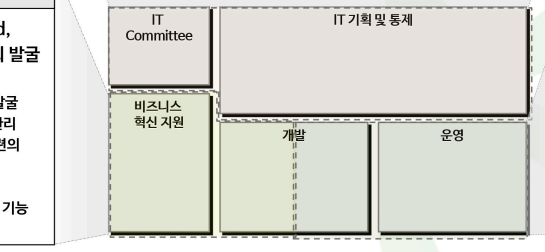
IT : 전사 IT Control Tower

Business와 연계된 IT 사업전략 수립(사업성 검토 포함)과 실현을 위한 두 개 Core에 대한 관리 및 모니터링 수행
→ IT팀의 IT R&D 및 전략 기능 강화
→ IT 표준화 및 기획/통제 프로세스 전반의 역할과 책임

혁신

기술 기반의 고객 및 기업 가치 제고

기술적 측면에서 Business Speed, Innovation 향상을 위한 개선 기회 발굴 및 혁신
→ IT ROI에 초점을 둔 IT 선도 요소 발굴
→ 지속적인 수요관리 및 Portfolio 관리
→ 요건정의, 설계, 개발에 이르는 일련의 프로세스의 유기적 연계
→ 사업부의 IT 투자결정에 기여하여 수익창출 역할을 수행하며, 사전에 기능 강화 및 체계 개선에 집중함



안정

IT 운영 최적화

Cost Ownership을 통해 운영비용의 절감 및 생산성 향상에 주력
→ 지속적으로 전사적 측면에서 IT 내부통제 및 문제점 개선
→ 매년 비용 절감을 주요 목표로 함
→ Transaction Volume 증가 등을 관리하기 위한 시스템 증설, 인프라 확충에 대한 비용 절감 노력
→ 개발 및 운영 인력 운영의 유연성을 확보하여 지속적인 생산성 향상

기존 IT 조직은 기술적 측면을 리딩하여 안정을 주도하고,
AI 등 신기술을 활용하는 IT 조직은 비즈니스 측면에서 혁신을 주도할 수 있도록 역할 조정

8

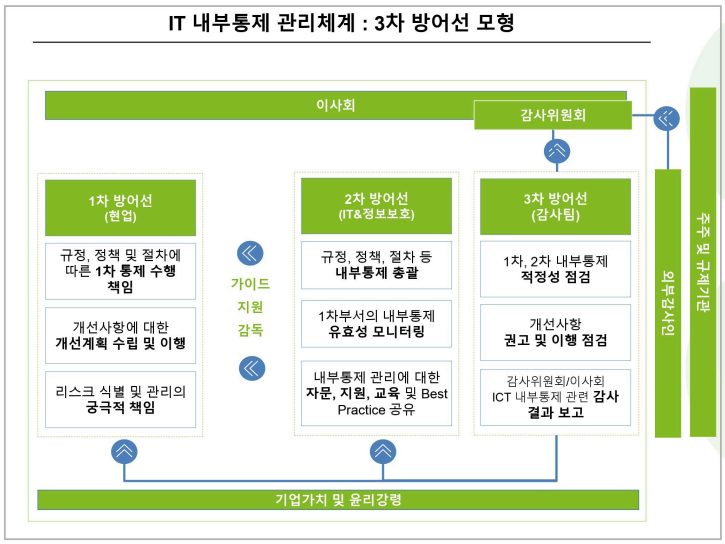
3.3. 주요 개선방안 – IT 구매 체계 일원화

업무 처리 효율화 ← Risk 관리 제고 → Trending

	분산적 (Decentralized)	분산된 거버넌스 (Distributed Governance)	중앙 주도형 (Center-Led)	중앙 집중형 (Centralized)
Description	• 소싱 및 구매 활동이 다양한 지역 및 / 또는 사업조직에서 발생함	• 각 기능 조직과 사업조직에서 구매와 소싱 활동에 대한 별도의 책임이 발생함	• 소싱 및 공급 업체 관리는 대부분의 카테고리에서 중앙 주도화 되어 있으며, 현지화 된 구매에만 국한적으로 분산되어 있음	• 모든 구매 및 전략적 소싱 활동 / 결정은 중앙 구매조직 내에서 이루어짐.
Benefits	• 현지(사업부별) 공급업체의 최대 활용 • 전략적 구매 / 소싱에 대한 결정은 사업부에 의해 이루어짐	• 부서 간의 공식적인 상호 작용 • 조직 전반에 걸쳐 구매를 활용할 수 있는 뛰어난 역량	• 비즈니스 영역 전문가는 여전히 요구 사항을 가지고 있지만 "비 핵심"구매 활동은 전문가에게 부담이 적어짐 • **공통성과 표준을 강화할 수 있는 능력**	• 부문간의 공통성/통일성 • 공급업체에 대한 공통된 접근법 • 공급 업체와의 협업에 집중 할 수 있음 • **품질 및 표준, 내부통제에 대한 향상된 기능**
Risks	• 가격 책정 / 대량 할인의 상호 보완적인 영향이 거의 없음. • 공급업체의 기반이 크고 관리가 어려움.(공식적인 계약 최소화) • 중복적인 일을 수행하게 될 가능성이 크다	• 구매/소싱에 대한 전담 조직이 없음 • 경영진 및 프로세스 간 의사 소통을 위한 효율적인 초점 없음 • 공통화/표준화 수준을 유지하기 어려움	• 소싱 전략 및 공급 업체 선택 사이의 충돌을 피하기 위해 구매와 비즈니스 영역 간의 목표 / 인센티브를 조정해야하며 적극적인 경영진의 지도력이 필요함	• 종종 큰 조직 변화가 필요함 • 운영 현장과의 유대가 약해짐 • 느린 응답 시간 • 더 적어지는 "현지 공급업체"

9

3.4. 주요 개선방안 – 충분한 IT 감사체계 마련

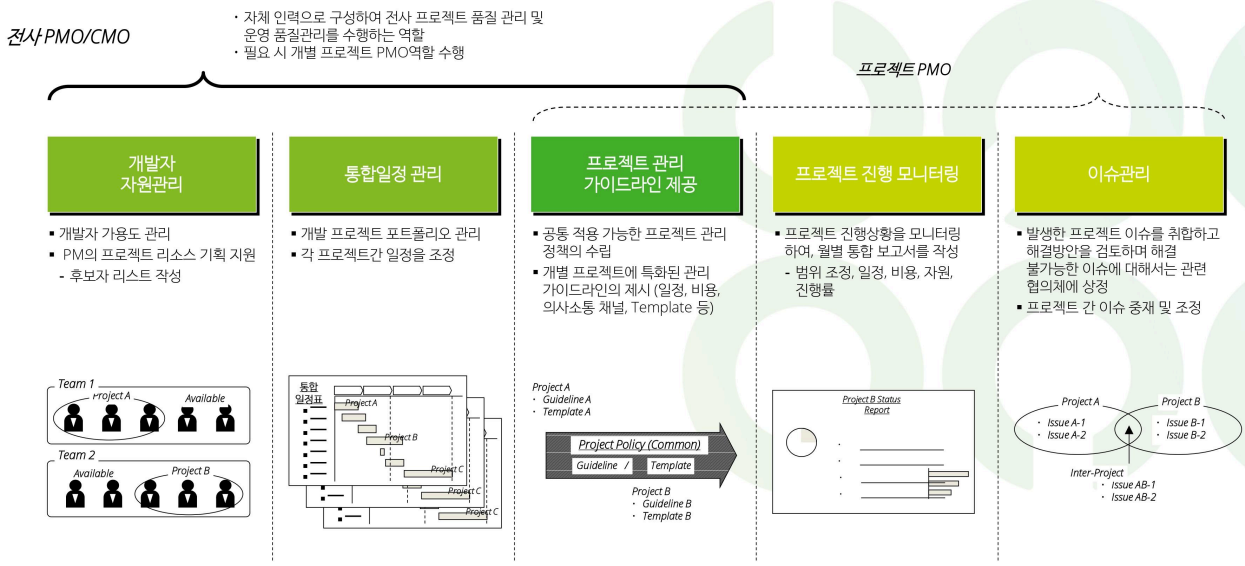


IT Risk 관리

IT 장애 예방 활동	• 외부 전문가를 활용하여 장애가 발생되는 원인에 대한 사전 진단
IT 비용 최적화 활동	• 필요 시 BMT, POC 수행을 통해서 최신 기술 검토하며, 벤더 종속성 탈피 노력 • Down Sizing 가능성 검토 뿐만 아니라 Cloud전환 등 지속적으로 수행 • 그룹 측면에서 비용 최적화 방안 제시
IT Risk 관리 강화	• IT운영 리스크 관리 체계, 지주사 IT진단 및 감사, 감사실 IT감사 강화, IT자체감사 및 품질 조직 강화
인식제고	• 내부통제는 형식적이고 불필요한 절차가 아닌 실질적으로 필요한 필수 활동으로 임직원 인식제고 • 룰 중심에서 원칙 중심의 Risk관리 강화

10

3.5. 주요 개선방안 - PMO/CMO 역할 강화



11

4. 결론



12

	Memo	

Track 2-5

최신 보안 위협 대응을 위해 무엇을 할 것인가?



박형근
시큐리티플러스 커뮤니티 리더

발표개요

최근 늘어가는 보안 사고 속에서 정보보호 관리체계의 무용론과 개선 요구가 높다. 점점 고도화되는 보안 위협에 대응하기 위해 보안 설계의 기본에 대해 되돌아 보는 시간을 가져 보고자 한다. 정보보호 관리체계에서 제로트러스트에 이르기까지, 특히 정보보안 프레임워크와 아키텍처에 대한 개념이 약한 많은 기업에게 많은 시사점과 통찰을 나누는 시간이 될 것이다.

이력

- 現 시큐리티플러스 커뮤니티 리더
- 前 IBM 보안 사업부 실장

최신 보안 위협 대응을 위해 무엇을 할 것인가?

박형근(securityplus@kakao.com)
시큐리티플러스 | 커뮤니티 대표

Confidential. For internal use only.

ISACA Korea Chapter

연사 소개

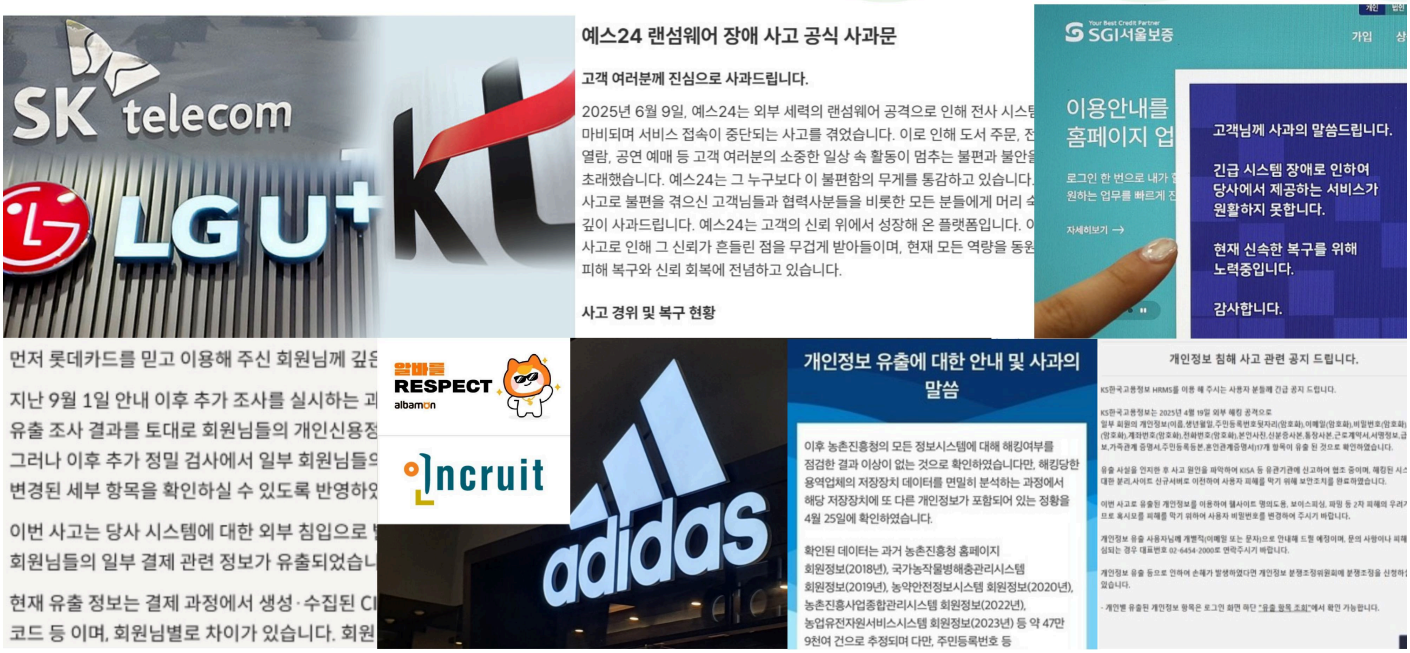


박 형 근
커뮤니티 대표
(securityplus@kakao.com)

- 現 정보보호 분야 NCS 확인 강사
- 現 시큐리티플러스 커뮤니티 대표(2004~현재)
- 現 IBM Rising Champions(2025.5~현재)
- 現 방산침해대응협의회 클라우드 자문위원, OWASP Seoul Chapter 정회원
- 現 정보통신기획평가원 및 중소기업기술개발 지원사업 평가위원
- 前 한국IBM 보안사업부 디지털 트러스트 부문 실장(2000.10~2025.4 정보보안 실무 경력 24년 7개월)
- 前 한국클라우드산업협회 클라우드 보안 분과위원회 위원
- 前 서울사이버대학교 정보통신학부 외래교수
- 前 방송통신위원회 미래전략 IT 자문 위원
- 前 국제 OASIS, ISF, TCG, FIDO Alliance, CSA, OpenGroup 정회원
- 前 CISSP, CISA, CGEIT



거듭되는 정보보안 사건 사고



최근 클라우드 보안 사고 사례

최근 클라우드 보안 사고는 인적 오류, 자격증명 관리 부실, 공급망 취약점 등 기초적 원인에서 비롯되어 대규모 데이터 유출 및 서비스 장애 등으로 이어짐.

1. TeleMessage(정부 메신저) AWS 서버 침해(2025.5)
   - 원인: 공개 서버 설정 오류, 관리자 인증정보 유출, 민감 데이터 암호화 미흡
   - 결과: 미국 정부 관계자 등 주요 인물 통신기록·연락처 공개, 백엔드 서버 공격 노출.

2. Oracle Cloud SSO/LDAP 데이터 유출(2025.3)
   - 원인: 서드파티(Third-party) 보안 취약점 및 클라우드 인프라 자격증명 관리 미숙, 인증/암호화 키 유출
   - 결과: 약 14만 명 테넌트의 암호화 키, 관리용 패스워드, 계정정보 유출, 다수 기업의 서비스 및 데이터 안전성 위협.

3. 스타벅스 써드파티 클라우드 공급망 침해(2024.11)
   - 원인: 공급망 연동사 클라우드 인증정보 유출, 업체간 권한 관리 미흡
   - 결과: 내부정보 및 직원 정보 일부 노출, 일시적 IT 장애와 조사·복구 비용 발생.

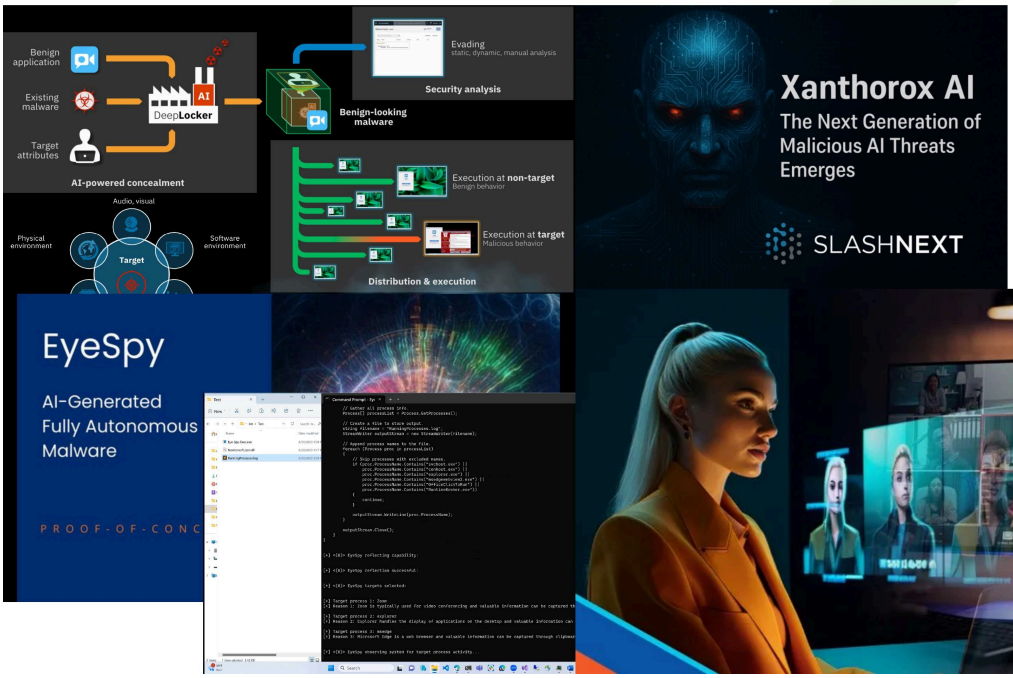
4. Snowflake 데이터 유출(2024)
   - 원인: 고객사 계정정보 탈취(InfoStealer, 피싱), MFA(다중 인증) 미적용, 오래된 사용자 자격증명 미삭제
   - 결과: Ticketmaster, Santander, AT&T 등 세계적 기업 포함 수백 곳 데이터 대량 유출, 민감정보 탈취, 평판 및 직접적·간접적 금전 손실, 피해기업별 법적·규제 문제 대두.

클라우드 보안 사고의 주요 원인

- ✓ 인적 오류 및 내부자 위협: 클라우드 설정 실수, 잘못된 공유 권한 설정, 약한 비밀번호 사용 등
- ✓ 접근 제어 및 아이덴티티 관리 부실: 다중 인증(MFA) 미도입, 접근권한 부적절 관리, 유출된 계정 혹은 접근 토큰(Access/Secret Key) 등
- ✓ 구성 오류 및 보안 업데이트 미비: 저장소(Bucket)나 데이터베이스가 인터넷에 공개되는 등 기본 설정 미흡, 중요 시스템에 최신 보안 패치 미적용 등
- ✓ 피싱 및 자격증명 탈취: 피싱 이메일을 통한 계정 탈취, 비밀번호 유출 후 불법 접속 시도
- ✓ 제3자 공급자 및 API 보안 취약: 협력사, 외부 연동 API 보안 취약점으로 인한 데이터 유출 및 시스템 침해 발생

5

인공지능 악용과 늘어가는 사이버 위협



6

최신 보안 위협 대응을 위해 무엇을 할 것인가?

공공 부문

범부처 정보보호 종합대책 발표(10.22) https://cafe.naver.com/securityplus/65347

주요 대책	고려 사항
주요 IT 자산에 대한 식별 관리 체계 (자산 관리) 구축	• 기업 거버넌스 하에서 각 부문별 정보보호 거버넌스와의 상호 유기적 협조 • 데이터 흐름 식별 통한 Shadow IT 탐지 및 관리
모의해킹 및 화이트 해커 활용하여 상시 보안 취약점 점검 체계 구축	• 비즈니스 우선 순위와 별도로 공격 표면 식별 및 관리 • 레드팀 운영을 포함하여 다양한 인공지능 기반 취약점 점검 기술 활용
정보보호 공시 의무 기업 확대	• E.S.G. 활동 하에서의 정보보호 주요 활동으로 함께 고려
다중 인증, 인공지능 기반 이상 탐지 체계	• 신뢰도 및 위험 기반 다중 인증 등 제로트러스트 아키텍처 기반 전환 • 인증 및 접근 로그와 인공지능 기반 이상 탐지 체계와의 연계
SBOM 제출 의무화	• CycloneDX 등 취약점 정보 포함한 SBOM 표준 활용 및 관리 • DevOps와의 통합 등 소프트웨어 라이프사이클 관리 프로세스 내 내재화

7

최신 보안 위협 대응을 위해 무엇을 할 것인가?

일반 부문

A 통신사 정보보호 혁신안(7.4)



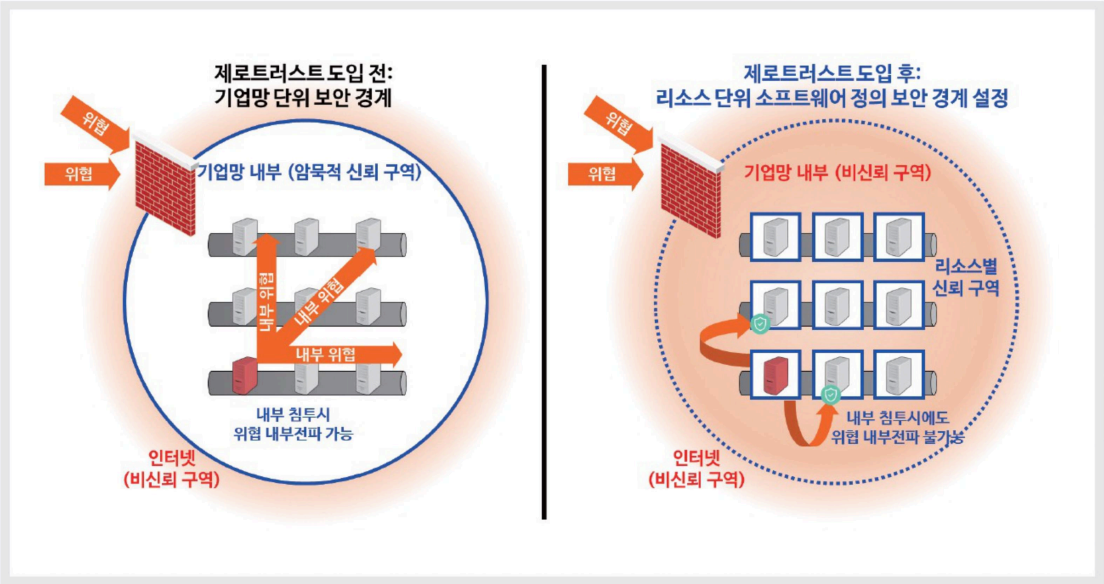
https://cafe.naver.com/securityplus/64606



https://cafe.naver.com/securityplus/65338

8

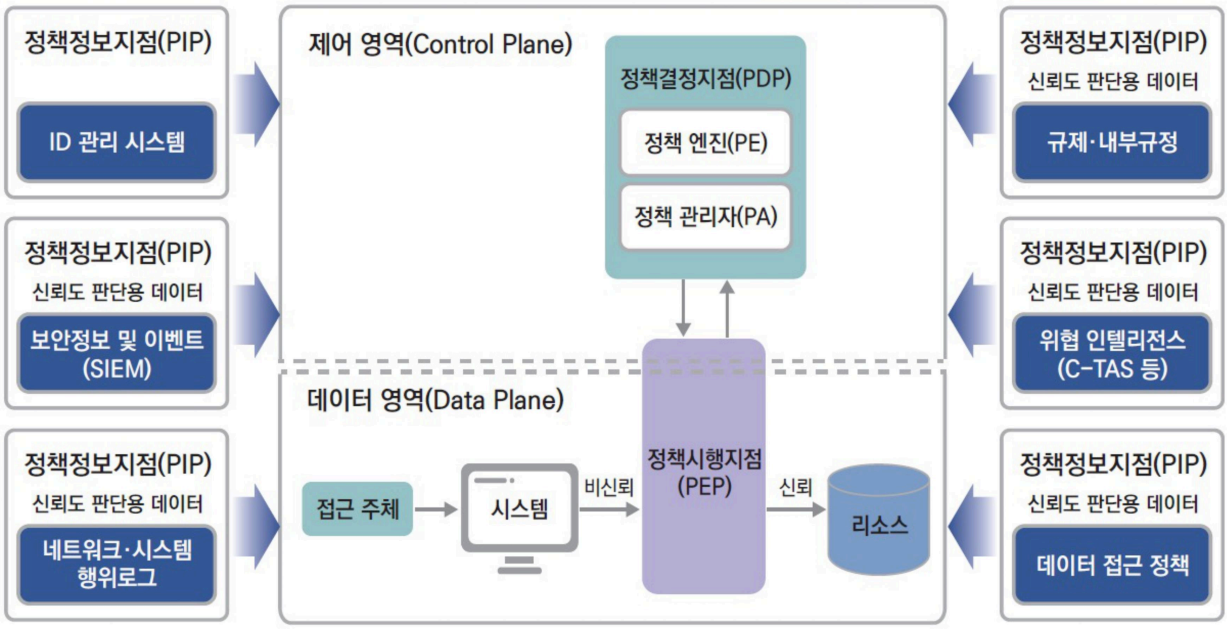
보안 패러다임의 변화



(출처: A.Kerman / NIST, "Zero Trust Cybersecurity:Never Trust, Always Verify"의 그림 재구성)

9

제로트러스트 아키텍처 보안 모델



10

제로트러스트 성숙도 모델 2.0에서 제시한 기능 및 세부역량

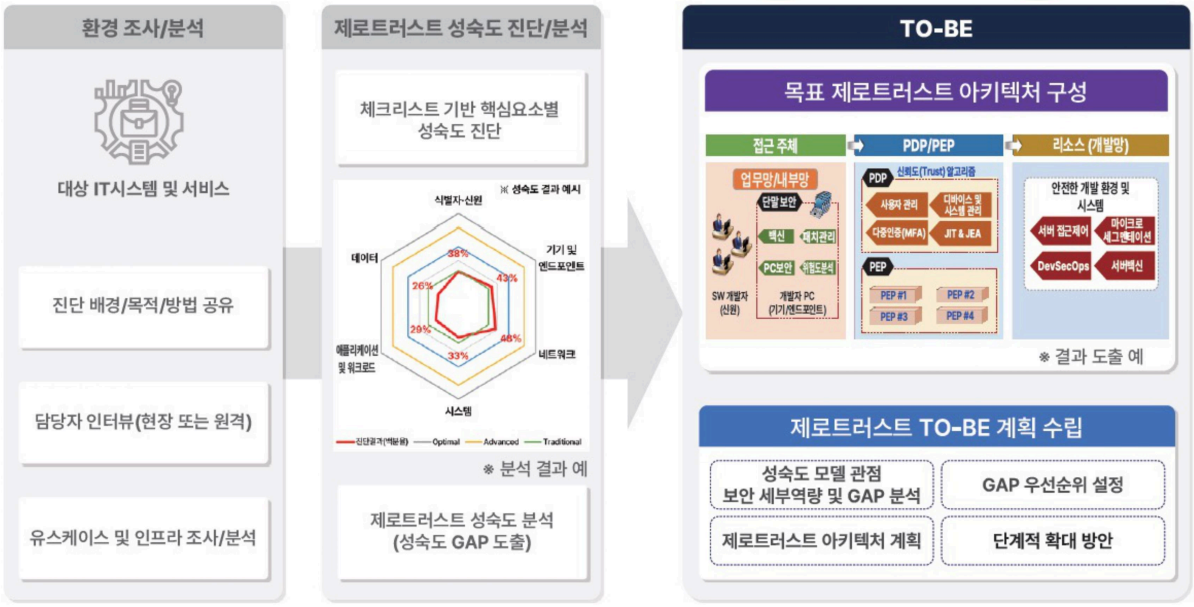


제로트러스트 성숙도 모델 4단계 수준

1. 기존 단계 (정적, 경계 기반, 수동)	2. 초기 단계 (일부 자동화)	3. 향상 단계 (자동화, 중앙집중적, 통합)	4. 최적화 단계 (동적, 완전 자동화)
▸ 주요 구성 요소들이 수동으로 설정되며, 정적인 보안 정책으로 인해 유연하지 못하게 정책시행 ▸ 경계 기반 보안 위주의 보안 아키텍처 구성 ▸ 수동으로 사고에 대응하며, 시스템에 대한 가시성이 제한적	▸ 일부 프로세스가 자동화되며, 핵심 요소별 연계가 일부 이루어짐 ▸ 속성 할당과 생명주기 관리가 부분적으로 자동화되며, 내부 시스템에 대한 기본적인 모니터링 제공 ▸ 프로비저닝 이후 최소 권한 변경에 대응 가능	▸ 자동화의 범위가 확장되고, 중앙 집중 제어가 강화되는 단계 ▸ 중앙 집중식으로 통합된 가시성 제공 ▸ 중앙 집중식 ID 관리를 통해 핵심 요소 간 상호작용에 기반한 정책 시행	▸ 자산 및 리소스에 대한 속성이 완전히 자동으로 할당되며, 동적인 정책이 적용되는 단계 ▸ 자동화된 트리거에 기반한 동적 정책 생성 ▸ 자산에 대해 동적 최소 권한 기반 접근 허용 ▸ 구성요소 간 상호운용성을 위한 개방형 표준 준수 이행 및 강화

12

제로트러스트 아키텍처 모델 구성 절차



13

준비 단계에서의 주의점



1. 자산(데이터, IT 자산 – 서버, 서비스 등)의 식별

* 기존 아키텍처 및 Shadow 인프라 고려
* 보안 문화적 접근 방법의 적극적인 도입

2. 객관적인 현재 보안 수준 진단

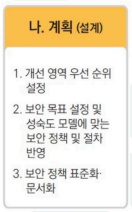
* 비즈니스 우선 순위 뿐만 아니라 공격 표면에 따른 위험도 반영

3. 기존 보안 아키텍처 및 기술 검증

* 정책 결정 지점(PDP), 정책 시행 지점(PEP), 와 정책 정보 지점(PIP)의 상호 작용 및 연계 가능 여부 확인 및 방향성(활용, 보완, 대체 등) 수립



계획 단계에서의 주의점



- ✓ **제로트러스트**는 "완전한 불신"이 아닌 "지능적 검증"을 구현하는 과정이며, 솔루션이 아닌 아키텍처의 전환입니다.
- ✓ **기술적 완성도보다 ▲조직 문화 정착 ▲점진적 확장 ▲교차 기능(자동화, 거버넌스) 통합 ▲종합적 모니터링**이 성패를 좌우합니다.
- ✓ 특히 내부자 위협과 공급망 공격에 대한 대비를 강화해야 합니다.

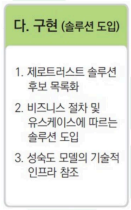
1. 로드맵과 성패 요인을 고려한 명확한 목표와 구체적 성과 지표 설정
* KPI 예시: 자동화 정책 실행율, 세션 재인증 빈도, 마이크로세그먼트 수 등

2. 제로트러스트를 위한 개념과 도입 기술의 이해, 보안 문화 및 커뮤니케이션 강화를 위해 직무와 직급에 따른 공통 혹은 심화 사용자 교육 및 변화 관리 계획 반영

3. 구현 단계의 기술적 고려사항을 계획 단계에 반영.

15

구현 단계에서의 주의점



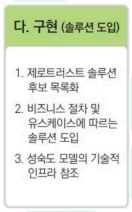
1. 시범 도입(Pilot) 필수: 소규모 환경(예: 특정 부서)에서 검증 없이 전사적 적용 시 레거시 시스템 충돌이 발생할 수 있음. 특히 클라우드-온프레미스 하이브리드 환경에서 호환성 테스트가 중요.

2. 정책 서버 보호: PDP/PEP를 겨냥한 DDoS 공격 시 전체 시스템 중단 위험이 존재. 이중화 및 오토스케일링(도커 및 쿠버네티스, 클라우드, SaaS 등) 적용 권장.

3. 신뢰도 평가 체계 구축: 단순 인증을 넘어 ▲기기 보안 상태(패치, 악성코드 검사 등) ▲위협 인텔리전스 ▲ 사용자 행동 분석(UEBA)을 통합한 동적 신뢰 평가 필요. 정적 정책만으로는 탈취된 권한을 가진 공격자를 탐지할 수 없음.

16

구현 단계에서의 주의점

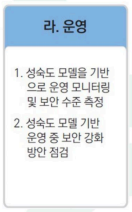


4. 세분화된 접근 제어(마이크로 세그멘테이션 전략): 네트워크 마이크로 세그멘테이션을 적용해 공격 표면을 최소화해야 함. 과도한 세분화는 관리 복잡성을 증가시킬 수 있으므로, 애플리케이션 중요도에 따른 계층화 중요.

5. 레거시 시스템 대응: SAML, OIDC(OpenID Connect), OAuth 2.0 등 상호 운용성 표준을 활용해 구축해야 함. 레거시 개조보다는 제로트러스트 전용 솔루션 도입 권고.

17

운영 단계에서의 주의점



1. 자동화 및 가시성: 인증 및 접근 통제 로그의 가시성 확보 중요. 지속적 모니터링으로 인한 로그 폭증은 가시성을 저하시킬 수 있으며, AI 기반 이상 탐지 도구로 대응 효율성을 높여야 함. 정책 실행의 80% 이상을 자동화해야 함.

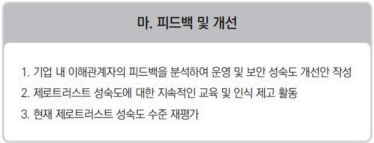
2. 보완 통제 활용: 과도한 차단 정책으로 인한 업무 지연을 방지하기 위해, 크리티컬 비즈니스 프로세스에 대한 예외 설정 시, 보완 통제를 반드시 고려해야 함.

3. 멀티 클라우드 환경 대응: 하이브리드 인프라에서의 일관된 정책 적용을 위해 클라우드 공급자별 API 통합 필수.

4. 내부자 위험 대비: 권한 탈취 시 피해 확산을 막기 위해 ▲세션 시간 단축(기본 10분 이하) ▲데이터 암호화 ▲비정상 행동 실시간 차단 적용.

18

지속적 개선 프로세스



1. 성숙도 모델 활용: NIST의 제로트러스트 성숙도 모델(ZTMM)을 기반으로 도입 단계별 진단 반복.

2. 실시간 모니터링 및 피드백: 사용자 행동 분석(UEBA)을 통해 비정상적 접근 패턴을 탐지하고, 인시던트 대응 계획을 정기적으로 개선. 정책 정보 지점으로 활용할 데이터 보완.

3. 성능 측정 및 피드백: 제로트러스트 침투 테스트를 정기적으로 수행해 효과성 분석. 특히 공급망 공격 시나리오(예: 악성코드 삽입 등)에 대한 대응력 평가.

4. 조직 문화 정착: 모든 임직원에게 "항상 검증" 원칙을 교육해 저항을 최소화해야 함. 월별 보안 인식 교육과 시뮬레이션 훈련 필수.

19

제로 트러스트에 대한 오해들

- 제로 트러스트는 기술 문제를 해결한다.
  - 제로 트러스트는 기술 문제를 해결하지 않습니다. 비즈니스 문제를 해결합니다.
- 제로 트러스트는 제품 또는 제품 집합이다
  - 제로 트러스트에 대한 일반적인 오해 중 하나는 ID 관리, 접근 제어 및 네트워크 세분화를 배포하면 제로 트러스트를 성공적으로 구현한 것입니다.
- 제로 트러스트는 직원을 신뢰하지 않는다는 것을 의미한다.
  - 제로 트러스트 접근 방식이 시스템을 신뢰할 수 있도록 만드는 것을 목표로 하지 않는다고 설명합니다.
- 제로 트러스트는 구현하기 어렵다
- '제로 트러스트 여정을 시작하는 올바른 방법'은 하나뿐입니다.
- SASE를 배포한다는 것은 제로 트러스트를 의미한다

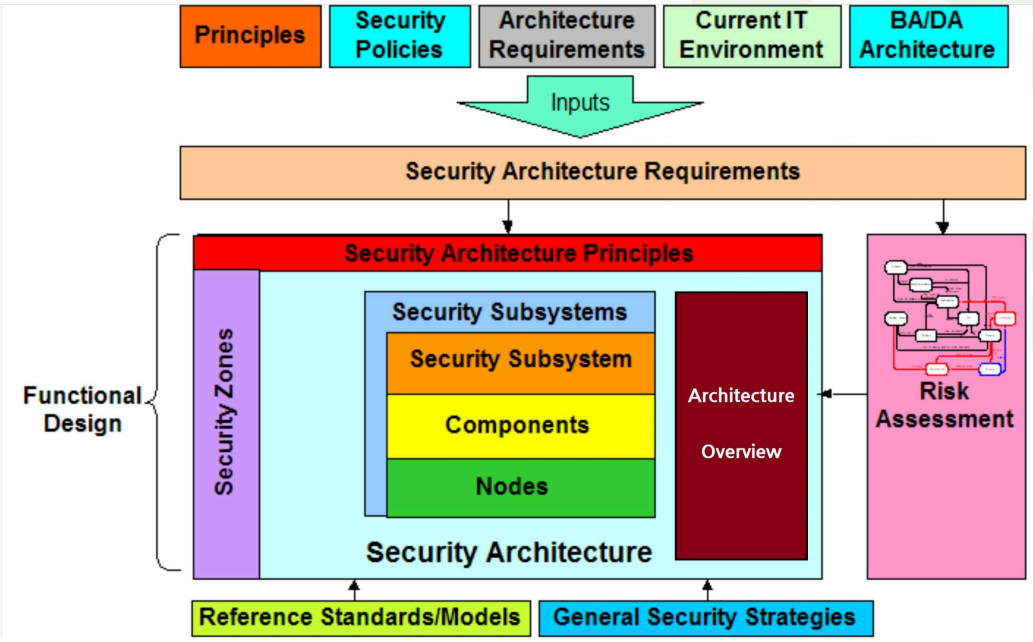
20

보안 아키텍처와 아키텍트가 요구되는 시대

기술, 사람, 프로세스를 포함한 기술을 사용하여 보안 문제를 고수준에서 분석하고 해결하는 것(사람).



보안 아키텍처 관점에서의 제로 트러스트



22

	Memo	

	Memo	

	Memo	

	Memo	

2025 공동 학술대회

Track 3

Track 3-1	Track 3-2	Track 3-3	Track 3-4	Track 3-5
AI 거버넌스와 조직·산업 혁신 좌장: 김남일 연세대학교 교수	AI와 신뢰·프라이버시 좌장: 이은곤 수원대학교 교수	AI와 인간 경험 좌장: 장재영 한국인터넷진흥원 연구위원	온라인 리뷰와 eWOM 좌장: 박하은 경희대학교 교수	가짜 리뷰 및 허위 정보 탐지 좌장: 박재영 국립부경대학교 교수
바른ICT연구소-전국정보보호정책협의회 MOU 협약식 김범수(연세대학교 바른ICT연구소) 김완집(전국정보보호정책협의회)	AI에 대한 인식이 규제 필요성과 수용 의도에 미치는 영향 김현정(연세대학교 바른ICT연구소)	디지털 혁신과 'Brain Rot'의 역설: 일본과 한국의 성인 역량 비교 분석 정미정(연세대학교 바른ICT연구소)	The Voice of the Player: Sentiment and Topic Shifts in Steam Game Reviews 김엘레나(순천향대학교), 호명휘(순천향대학교), 최재원(순천향대학교)	A Hybrid Multi-Perspective Approach to Fake Review Detection 사비타(순천향대학교), 가중정(순천향대학교), 최재원(순천향대학교)
Agentic AI 보안 위협 동향 조효진(연세대학교 정보대학원)	중소기업의 스마트 제조 도입 영향 분석 김민호(연세대학교 바른ICT연구소)	자동화 시대의 일과 삶: 한국 경제의 증거를 중심으로 한 근로자 웰빙 연구 김미경(연세대학교 바른ICT연구소)	The Impact of Text and Image Clues on Consumer Trust and Purchase Intention in the eWOM Situation 진우교(순천향대학교), 유여완(순천향대학교), 최재원(순천향대학교)	Extreme and Deceptive Reviews in Hospitality: Evidence from YouTube Content 부니요드(순천향대학교), 가중정(순천향대학교), 최재원(순천향대학교)
Scale-Free 네트워크와 Small-world 네트워크 허위정보 확산 및 억제전략 설계 노병규(연세대학교 바른ICT연구소)	AI 활용 전략이 직무 성과에 미치는 영향 천정은(연세대학교 바른ICT연구소)	생성형 AI의 어두운 그림자 이준혁(연세대학교 바른ICT연구소)	When Extremes Deceive: The Role of Hedonic and Utilitarian Motivations in Perceiving Manipulation in Online Reviews 이맹릉(순천향대학교), 호명휘(순천향대학교), 최재원(순천향대학교)	보이스피싱 탐지를 위한 이론적 접근 김재희(연세대학교), 권성훈(연세대학교), 김희웅(연세대학교)
Evaluation and Rule-Making of Agentic AI 김범수(연세대학교 바른ICT연구소), 장백철(연세대학교 바른ICT연구소), Jason G. Allen (Centre for Digital Law, SMU)	AI 리터러시를 통한 AI 종차별주의의 부정적 영향 완화 기제 탐색 노환호(연세대학교 바른ICT연구소)			

	Memo	

Track
3-1

AI 거버넌스와 조직·산업 혁신

좌장: 김남일 연세대학교 교수

바른ICT연구소-전국정보보호정책협의회 MOU 협약식

김범수(연세대학교 바른ICT연구소)
김완집(전국정보보호정책협의회)

Agentic AI 보안 위협 동향

조효진(연세대학교 정보대학원)

Scale-Free 네트워크와 Small-World 네트워크 허위정보 확산 및 억제전략 설계

노병규(연세대학교 바른ICT연구소)

Evaluation and Rule-Making of Agentic AI

김범수(연세대학교 바른ICT연구소)
장백철(연세대학교 바른ICT연구소)
Jason G. Allen(Centre for Digital Law, SMU)

Agentic AI 보안 위협 동향

조효진
연세대학교 정보대학원 교수

Agentic AI는 인간의 명시적 지시 없이도 목표를 스스로 설정하고, 외부 환경을 인식하며, 행동 계획을 수립·수정하는 자율형 인공지능으로, 단순한 생성형 AI를 넘어 행동하는 AI로 진화하고 있다. 최근 Anthropic에서는 자율적 의사결정 루프를 포함한 MCP (Model Context Protocol) 아키텍처를 도입하면서, AI가 외부 도구와 직접 상호작용하는 환경이 빠르게 확산되고 있다. 그러나 이러한 자율성은 보안 측면에서 새로운 공격 표면을 만들어낸다.

본 연구는 MCP 환경에서 발생할 수 있는 공급망 공격, 툴 하이재킹 공격, 프롬프트 공격 등의 시나리오를 소개하고 실제 공격 과정을 시연한다. 또한, 이러한 위협에 대응하기 위한 Agentic AI 보안 기술 동향을 간략히 정리함으로써, 향후 자율형 AI의 안전한 활용과 보안 프레임워크 구축 방향 설정에 기여한다.

KEYWORDS Agentic AI, Model Context Protocol, Tool Hijacking, Supply Chain Attack, Prompt Injection, AI Security

Scale-Free 네트워크와 Small-world 네트워크 허위정보 확산 및 억제전략 설계

노병규
연세대학교 바른ICT연구소 연구교수

디지털 플랫폼의 발전은 정보 확산 속도를 비약적으로 증가시켰으나, 동시에 허위정보(misinformation)의 확산이라는 심각한 사회적 위협을 야기하고 있다. 허위정보의 확산은 사회에 대한 신뢰 저하와 분열을 심화시켜 사회 전반의 안정성을 위협하는 구조적 문제이다.

본 연구에서는 대표적인 네트워크 모델인 Scale-Free 네트워크와 Small-world 네트워크를 대상으로 정보 확산 양상을 비교 분석하고 최적의 대응 방안을 설계한다. 이를 위해 5가지 허위정보 억제 전략을 설계하였다. 각 전략은 1,000개의 노드를 가지는 두 네트워크 모델에서 SIR 모델을 활용한 허위정보와 진실정보의 경쟁적 확산 시뮬레이션을 통해 검증하였다. 또한 각 전략을 허위정보 확산 단계(발생, 초기 확산, 급속 확산, 확산 정점, 안정 단계)에 적용하고 단계별로 최적 조합 전략을 제안한다.

본 연구는 기존의 사후적 차단 접근에서 벗어나 선제적 진실정보 확산 전략의 효과성을 네트워크 모델과 대응 전략별로 실증하였으며, 허위정보 확산 상황에 따른 차별화된 대응 가이드라인을 제공한다. 이를 통해 네트워크 특성을 고려한 맞춤형 허위정보 대응 체계 수립에 기여할 것으로 기대된다.

Evaluation and Rule-Making of Agentic AI*

김범수
연세대학교 정보대학원 교수, 바른ICT연구소 소장
–
장백철
연세대학교 정보대학원 교수, 바른ICT연구소 부소장
–
Jason G.Allen
Director, Centre for Digital Law, Singapore Management University

Agentic AI는 인간의 명시적 지시 없이도 스스로 목표를 설정하고, 환경을 인식하며, 실행 전략을 조정하는 자율적 인공지능으로, 기존 생성형 AI를 넘어서는 새로운 기술적 패러다임으로 부상하고 있다. 이러한 자율성은 AI가 상황에 따라 스스로 규칙을 생성하고 판단을 내리는 능력을 가능하게 하지만, 동시에 기존의 성능 평가 기준과 규제 체계로는 그 행동을 검증하기 어려운 문제를 야기한다. 최근 주요 글로벌 AI 모델을 대상으로 한 실증 분석 결과, Agentic AI의 의사결정 루프는 인간 개입 수준이 낮아질수록 예측 불가능성이 증가하고, 동일 입력 대비 결과 편차가 평균 28~47%까지 확장되는 것으로 나타났다. 이는 기존의 정태적 성능 평가 지표로는 AI의 자율적 판단 과정을 충분히 포착할 수 없음을 시사한다.

본 연구는 Agentic AI의 자율적 행동을 데이터 기반으로 측정하고, 평가와 규칙 생성 간의 동적 균형을 실증적으로 규명함으로써, 향후 AI 평가 프레임워크의 진화 방향과 자율 AI의 신뢰성 확보를 위한 정량적 기준 수립에 기여한다.

KEYWORDS Agentic AI, Autonomous Decision-Making, Evaluation Framework, Decision Traceability Index

*이 발표는 김범수, 장백철, Jason G. Allen의 공동 연구를 기반으로 하며, 이 중 일부 내용은 해당 공동 연구에서 발췌 및 재구성하여 사용되었다. 본 연구는 대한민국 교육부와 한국연구재단의 지원(NRF-2024S1A5C3A03046579)을 받아 수행된 연구임

Track
3-2

AI와 신뢰·프라이버시

좌장: 이은곤 수원대학교 교수

AI에 대한 인식이 규제 필요성과 수용 의도에 미치는 영향

김현정(연세대학교 바른ICT연구소)

중소기업의 스마트 제조 도입 영향 분석

김민호(연세대학교 바른ICT연구소)

AI 활용 전략이 직무 성과에 미치는 영향

천정은(연세대학교 바른ICT연구소)

AI 리터러시를 통한 AI 종차별주의의 부정적 영향 완화 기제 탐색

노환호(연세대학교 바른ICT연구소)

생성형 AI에 대한 인식이 규제 필요성과 수용 의도에 미치는 영향

김현정
연세대학교 바른ICT연구소

본 연구는 생성형 인공지능(Generative Artificial Intelligence, 이하 생성형 AI)에 대한 시민의 인식이 어떻게 규제 요구(regulatory preference)와 수용 의도(adoption intention)라는 상이하면서도 상호보완적인 태도로 형성되는지를 규명하고자 한다. 생성형 AI는 의료, 금융, 채용 등 고위험(high-risk) 영역에서 혁신성과 효율성을 제공하는 동시에, 불투명성, 책임소재의 불명확성, 편향, 프라이버시 침해 등의 위험을 동반한다. 이에 일반 대중은 AI를 기회(opportunity)와 위협(threat)으로 동시에 인식하며, 이러한 양가적 태도는 신뢰(trust)를 매개로 정책적 판단으로 전이된다. 본 연구는 기술수용모형(TAM), 위험–편익 이중경로이론(Dual Pathway of Risk and Benefit Perception), 다차원 신뢰이론(Multi-dimensional Trust Theory), 문화 차원 이론(Hofstede's Model of National Culture)과 주관적 사회경제적 지위(Subjective Socioeconomic Status)를 통합하여, 기회·위협 인식이 인지적(cognitive)·정서적(affective)·제도적(institutional) 신뢰를 경유해 규제 요구와 수용 의도로 이어지는 통합적 연구모형을 제시한다. 특히 제도적 신뢰를 '혁신을 억제하는 대체재'가 아닌, 신뢰를 강화하여 수용을 촉진하는 '보완적 규제 메커니즘'으로 개념화하였다.

본 연구는 한국, 싱가포르, 미국의 일반 대중을 대상으로 의료·금융·고용 분야의 생성형 AI 활용 시나리오를 구성한 비네트 기반 반복측정 실험을 실시하며, 설명가능성, 책임성, 데이터 투명성 수준을 조작하여 고신뢰·저신뢰 조건을 설계한다. 응답 결과는 다층 회귀분석(multi-level regression analysis)을 통해 분석한다. 시나리오(Level 1)와 응답자(Level 2)의 중첩 구조를 고려하여 랜덤효과를 추정하고, 부트스트랩 기반 매개효과 검증과 교차수준 상호작용(cross-level interaction) 분석을 병행한다. 분석에는 Stata 18을 사용하며, 국가 간 비교를 위한 다집단 분석과 모형 적합도(AIC, BIC, ICC, LR test)를 통해 내·외적 타당성을 확보한다. 본 연구는 반복측정 실험과 다층 분석을 결합하여 생성형 AI 인식의 개인 내 및 사회문화적 요인을 통합적으로 규명하고자 하며, 이를 통해 생성형 AI 거버넌스 연구에서 신뢰 기반 규제(trust-based regulation)의 이론적 근거를 제시함으로써 시민의 사회문화적 맥락과 심리적 요인을 통합적으로 반영한 책임 있는 AI 수용(responsible AI adoption) 모델을 제안한다.

KEYWORDS 생성형 AI 기회인식, 생성형 AI 위협 인식, 규제 요구, 수용 의도, 문화적 가치, 주관적 사회경제적 지위

Analysis of Digital Transformation Effects in SMEs: Causal Inference Using PSM-DID during the COVID-19 Period

Minho Kim
Yonsei Univ. Barun ICT Research Center

This study examines the causal effects of digital transformation on small and medium-sized enterprise (SME) performance during the COVID-19 pandemic. Using data from the 2022 Korean Innovation Survey (KIS), we analyze 1,139 SMEs observed in 2019 (pre-COVID) and 2021 (post-COVID).

To address selection bias, we employ propensity score matching (PSM) using seven firm characteristics, obtaining 405 matched pairs through nearest neighbor matching. Post-matching diagnostics confirm excellent covariate balance, with all standardized mean differences below 0.1. We then apply difference-in-differences (DID) with two-way fixed effects to control for unobserved heterogeneity and common time trends.

Our results show that digitally transformed SMEs achieved 7.06% higher sales growth relative to non-adopters during the pandemic period ($p=0.003$). This effect is substantially larger than the naive DID estimate without matching (2.21%, $p=0.410$), highlighting the importance of addressing selection bias. The one-way firm fixed effects model yields 10.44% ($p<0.001$), while adding year fixed effects reduces this to 7.06%, indicating that approximately 3-4 percentage points reflect common recovery trends rather than transformation effects.

Heterogeneity analyses reveal that younger firms (≤21 years) experience an 8.95% effect compared to 1.97% for older firms, though subsample sizes limit statistical power. No significant heterogeneity emerges by export orientation or industry.

This study demonstrates the importance of combining PSM with DID for credible causal inference and provides robust evidence that digital transformation enhanced SME resilience during the pandemic. Our findings support targeted digital transformation policies, particularly for younger firms. Key limitations include the inability to test pre-trends and the focus on short-term effects, suggesting the need for longer panel analyses.

KEYWORDS Digital transformation, Small and medium-sized enterprises, Propensity score matching, Difference-in-differences, Fixed effects, Causal inference, COVID-19

AI 활용 전략이 직무 성과에 미치는 영향

천정은
연세대학교 바른ICT연구소 연구교수

최근 생성형 인공지능(Generative AI)의 급속한 확산은 개인의 업무 방식과 성과에 중대한 변화를 가져오고 있다. 본 연구는 직장인의 생성형 AI 활용 전략이 AI 활용 능력과 직무 성과에 미치는 종단적 관계를 규명하고자 하였다. 특히, 문제 해결 중심의 활용 전략과 회피 중심의 활용 전략이 생성형 AI 활용 능력을 매개로 직무 성과에 미치는 차별적 영향 경로를 분석하였다. 이를 위해 400명의 직장인을 대상으로 2025년 5월(1차)과 6월(2차), 약 한 달 간격으로 동일한 문항(대처 전략, AI 활용 능력, 직무 성과)을 두 차례 측정하였다. 수집된 자료는 교차지연 패널모형(CLPM) 및 반종단적 매개효과 분석(half-longitudinal mediation analysis)을 통해 분석하였다. 분석 결과, 문제 해결 전략은 생성형 AI 활용 능력을 향상시켜 직무 성과를 높이는 긍정적 경로를 보인 반면, 회피 전략은 AI 활용 능력을 저해하여 직무 성과에 부정적 영향을 미치는 것으로 나타났다. 본 연구는 생성형 AI 시대에 개인의 활용 전략이 기술 활용 역량과 업무성과에 어떻게 연결되는지를 규명함으로써, 조직 차원의 AI 적응 교육 및 인적자원 관리 전략에 실질적 시사점을 제공한다.

AI 리터러시를 통한 AI 종차별주의의 부정적 영향 완화 기제 탐색

노환호
연세대학교 바른ICT연구소 연구교수

AI와 인간의 교류는 점점 가속화될 것이다. 하지만 모방이나 통계적 추론은 가능해도 AI 입장에서 인간에 대해 제대로 “이해”하기 어려운 특성들이 있을 것으로 보인다. 대표적으로, 왜 인간은 행복과 같은 감정적 경험을 그토록 중시할까? 강의를 통해 최근의 진화심리학 관점으로 재해석하는 행복(감정)의 기능과 존재 이유를 소개한다. AI와 더불어 사는 시대에 여러 기술적, 철학적 함의점이 담긴 질문이다.

AI와 인간의 교류는 점점 가속화될 것이다. 하지만 모방이나 통계적 추론은 가능해도 AI 입장에서 인간에 대해 제대로 “이해”하기 어려운 특성들이 있을 것으로 보인다. 대표적으로, 왜 인간은 행복과 같은 감정적 경험을 그토록 중시할까? 강의를 통해 최근의 진화심리학 관점으로 재해석하는 행복(감정)의 기능과 존재 이유를 소개한다. AI와 더불어 사는 시대에 여러 기술적, 철학적 함의점이 담긴 질문이다.

KEYWORDS 스마트폰

	Memo	

Track
3-3

AI와 인간 경험

좌장: 장재영 한국인터넷진흥원 연구위원

디지털 혁신과 'Brain Rot'의 역설:
일본과 한국의 성인 역량 비교 분석

정미정(연세대학교 바른ICT연구소)

자동화 시대의 일과 삶:
한국 경제의 증거를 중심으로 한 근로자 웰빙 연구

김미경(연세대학교 바른ICT연구소)

생성형 AI의 어두운 그림자

이준혁(연세대학교 바른ICT연구소)

디지털 혁신과 'Brain Rot' (뇌썩음)의 역설: 일본과 한국의 성인 문해력 비교 분석

정미정
연구교수, 연세대학교 바른ICT연구소

한국은 세계 최고 수준의 디지털 인프라를 갖추고 있음에도 불구하고 성인 문해력이 급격히 하락하는 '디지털 역설' 현상을 보이고 있다. 경제협력개발기구(OECD)의 국제 성인 역량 조사(PIAAC) 1주기(2012)와 2주기(2023) 데이터를 비교하면, 한국 성인의 역량 저하 추세는 뚜렷하게 나타난다. 이러한 문해력 저하 현상은 연령 증가에 따른 역량 감소 폭이 OECD 국가 중 가장 크다는 점에서도 심각성을 더하며, 중장년층과 노년층은 급격한 저하를 보여 사회 통합과 경제적 형평성에 위험을 초래하고 있다. 특히, 유사한 디지털 환경 속에서도 높은 역량 수준을 유지하고 있는 일본과 극명한 대조를 보였다. 본 연구는 PIAAC 마이크로 데이터에 대한 다중 회귀 분석과 양국의 평생 학습 정책에 대한 비교 분석을 통해, 이러한 격차를 매개하는 구조적 요인을 규명한다.

연구 결과, 디지털 기술 활용은 이중적인 영향을 미치는 것으로 나타났다. 정보를 능동적으로 탐색하는 디지털 활동 (능동적 ICT 활용)은 양국의 성인 문해력과 긍정적인 상관관계를 보인 반면, 수동적이고 오락 중심적인 콘텐츠 소비 (수동적 ICT 활용)는 성인 문해력에 긍정적인 영향을 미치지 못하거나 오히려 부정적인 영향을 미쳐, 미묘한 '뇌썩음(brain rot)' 가설을 뒷받침한다. 결정적으로 본 연구는 양국의 상이한 평생 학습 정책 철학이 이러한 격차를 유발하는 핵심적인 구조적 변수임을 실증한다. 인적 자본론(human capital theory)에 크게 의존하는 한국의 정책 프레임워크는 경제적 성과와 특정 직무 기술 향상에 초점을 맞춘 반면, 전후(戰後) '사회 교육법'에 기반하고 공민관(公民館, Kominkan)이라는 지역 학습 센터를 통해 제도화된 일본의 성인 학습 모델은 전인적 발달과 사회적 자본 형성을 촉진한다.

본고는 양국의 성인 문해력 격차가 개인적 선택의 총합이 아닌, 정책이 설계한 시스템적 결과라고 주장한다. 따라서 한국이 성인 문해력의 구조적 하락을 해결하기 위해서는 현재의 경제적 성과 및 직무 기술 중심의 평생학습 정책을 넘어서, 일본의 공민관(Kominkan) 사례처럼 전인적 발달과 사회적 자본 형성을 지원하는 지역사회 기반의 '인지적 완충 장치 (cognitive buffer)' 혹은 '인지적 예비능 (cognitive reserve)'을 제도적으로 구축하는 방향으로 평생학습 정책 철학을 확장해야 한다고 제안한다.

KEYWORDS Digital Paradox, Brain Rot, 성인 문해력, PIAAC, 평생학습정책

자동화 시대의 일과 삶:
한국 경제의 증거를 중심으로 한 근로자 웰빙 연구



김미경
(연세대학교 바른ICT연구소 연구교수)

2025. 11. 13
Barun ICT Research Conference
한국지식경영학회 추계학술대회
ISACA Korea Conference

Introduction: The Automation Era

The Global Shift
Rapid integration of robots and automated systems is f
undamentally transforming workplaces worldwide.

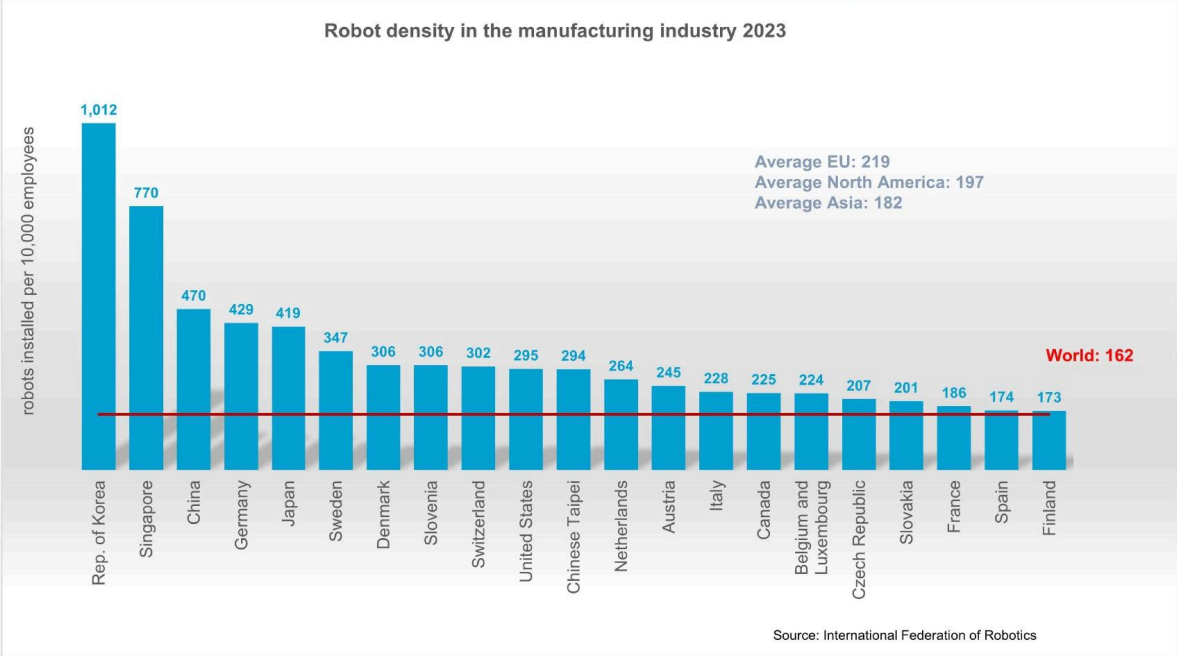
Productivity vs. People
While automation drives productivity and economic growth, i
t also introduces unprecedented challenges for workers.

The Research Gap
Extensive research on economic consequences (job displacement, wages) but a critical gap in understanding psychological impacts.

Our Contribution
Investigating workers' experiences with industrial automation through the lens of Technostress Theory.

Made with GAMMA

Why South Korea? A Unique Case Study



Made with GAMMA

AI and Robotics Impact on the Labor Market

- **Frey-Osborne (2017)** method is used to estimate occupational replacement probability.
- **Acemoglu-Restrepo (2022)** method is used for a regional labor market approach.
- **IFR (International Federation of Robotics)** data provides the number of industrial robots at the country level.
- **Bartik** instruments are used to control for endogeneity by correlating robot density with a region's initial industrial structure. This captures changes across all industries, regardless of country.
- The general consensus is that increased robot adoption reduces employment.

Made with GAMMA

Recent Studies: Moving Beyond the Regional/Industrial Level to Analyze the Impact on Individuals.

•**Giuntella et al. (2023)**: A study on China found that as robot exposure increases, so does the likelihood of quitting a job, receiving job training, retiring, and spending time with family.

•**Anelli et al. (2024)**: Found that the value of unmarried men in the marriage market decreased due to robot adoption, leading to negative effects on marriage and fertility.

•**Anelli et al. (2021), Im et al. (2019)**: Found that an increase in robot exposure in Europe led to an increase in support for far-right and populist parties; **Frey et al. (2018)** confirmed the correlation with Trump's election.

•**Thewissen and Rueda (2019)**: Found that an increase in robot exposure in Europe led to a greater preference for redistribution.

- **In addition, automation can elevate workplace stress, exacerbate job insecurity, and reduce workers' sense of control (Nazareno and Schiff, 2021; Abeliansky, Ahammer, and Lenk, 2024).**

Made with GAMMA

Technostress Theory: Extending the Framework

Originally developed for psychological strain from digital technology
(e.g., constant emails, mobile notifications).

Our research extends technostress to industrial automation – where its relevance is increasingly apparent but understudied.

Beyond the Shop Floor: Even factory robots c an reshape after-hours mindsets through heightened performance quotas and job-security anxieties, 'invading' workers' off-shift mental space.



Key Dimensions of Technostress in Automation

Techno-Overload
Pressure to work faster and longer with automated systems

Techno-Invasion
Blurring of boundaries between work and personal life (through persistent cognitive burdens)

Techno-Complexity
Difficulty in understanding and a dapting to new robotic t echnologies

Techno-Insecurity
Fears about job security and skill obsolescence due to automation

Techno-Uncertainty
Anxiety about constant technological changes and ambiguous future roles

Made with GAMMA

Research Questions & Hypotheses

1

RQ1: Mental Health
Does robot adoption affect stress, depression, and coping behaviors?

- **H1: Robotization ↑ → Stress ↑**
- **H2: Robotization ↑ → Depression ↑**
- **H3: Robotization ↑ → Alcohol Consumptio n ↑**

2

RQ2: Job Satisfaction
Does robot adoption influence overall and specific d imensions of job satisfaction?

- **H4: Robotization ↑ → Overall Job Satisfaction ↓**

Made with GAMMA

Empirical Strategy: Causal Identification

Core Approach

Two-way fixed-effects instrumental-variables (IV) models

Data Linkage

Connecting county-level robot-intensity measures to individual-level outcomes

Time Horizons

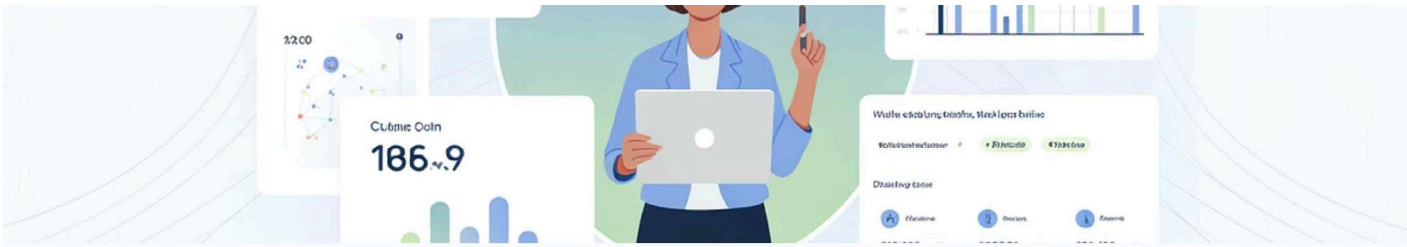
Analyzing effects over one-, two-, and three-year intervals to capture immediate and gradual impacts

Addressing Endogeneity

Constructing a Bartik-style instrument leveraging foreign robot adoption trends (Singapore, Taiwan, Germany, Japan, China) at the industry level

- **Relevance:** Foreign robotization correlates with domestic adoption
- **Excludability:** Foreign trends affect Korean outcomes only through domestic robot penetration

Made with GAMMA



Data Sources

Community Health Survey (CHS)

- Nationally representative, annual survey
- **Outcomes: Perceived stress, Depressive symptoms, Self-rated health, Monthly alcohol consumption (men)**
- Strength: Rich psychological and behavioral outcomes, regionally disaggregated

Korean Labor and Income Panel Study (KLIPS)

- Longitudinal survey of urban households
- **Outcomes: Overall job satisfaction and 9-subcomponents (e.g., compensation, stability, task content, meaning, career development)**
- Strength: Workplace-specific outcomes, within-individual changes over time

Made with GAMMA

Table 1. Summary Statistics of Dependent Variables (CHS)

Variables	Mean	SD	Min	Max	Definition
Perceived Subjective Health Level*	45.55	6.61	29.5	74.9	People who reported subjective health as "very good" or "good"/survey respondents x 100
Perceived Stress*	26.75	4.68	3.7	40.7	People who responded that they feel "a lot" or "a fair amount" of stress in their daily lives / surveyed x 100
Depression experience rate*	5.77	2.39	0.1	17.7	People who have experienced feelings of sadness or hopelessness that interfered with their daily life for two or more consecutive weeks in the past year/survey respondents x 100
Monthly drinking rate (Male)*	72.94	4.98	54.8	85.7	Percentage of people who drank alcohol at least once a month in the past year.
Observations	1929				

Source: Community Health Survey (Korea Centers for Disease Control and Prevention)

* Age-standardized to the 2005 population estimate

Made with GAMMA

We follow **Giuntella et al. (2024)** in linking city-level robot exposure to individual outcomes, and we mirror **Abeliansky, Beulmann & Prettner (2024)** in applying a panel IV specification to isolate exogenous variation in regional robot intensity.

$$y_{ijt} = \beta_0 + \beta_1 Robot_{jt} + \Omega\Theta_{ijt} + \lambda_i + \eta_t + \epsilon_{ijt}$$

where $Robot_{jt}$ is robot exposure of region (city and county) j at time t, and β_1 measures individual's various subjective well-being. Θ_{ijt} is a vector of various individual-level covariates, such as gender, education, and industry.

$$Robot_{j,t} = \sum_s l_{s,t_0} \left(\frac{R_{s,t}}{L_{s,t_0}} \right), \text{ with } S = 19$$

where R_{st} is the number of robots in industry s at time t

L_{s,t_0} is employment in industry s in the baseline year 2000
l_{s,t_0} is the share of employment in industry s in region j as of 2000.

$$\boldsymbol{Robot_{jt}^{IV} = \sum_s l_{s,t_0} \left(\frac{R_{s,t}^{foreign}}{L_{s,t_0}} \right), \text{ with } S = 19}$$

Made with GAMMA

Table 2. Countries Included in the Bartik Variable Analysis

Bartik variable (IV)	Countries
IV1	Singapore
IV2	Singapore, Taiwan, Germany, and Japan
IV3	Singapore, Taiwan, Germany, Japan, and China
IV4	Finland, France, Italy, Sweden, Spain, United Kingdom, and Germany from Dauth et al. (2021)

Made with GAMMA

Key Findings: Mental Health & Job Satisfaction

Mental Health (CHS Data)

Outcome (2-Year)	Coefficient	Significance
Perceived Stress	+0.188	**
Depression	+0.122	***
Monthly Drinking	+0.210	***
Self-Rated Health	-0.619	***

Effects consistent across 1, 2, and 3-year intervals, aligning with technostress theory

Job Satisfaction (KLIPS Data)

Overall: Broad job-satisfaction ratings remain largely unchanged

Nuanced Erosion (Staggered Pattern):

- **1-Year:** Significant decline in Enjoyment of Daily Duties (reflects techno-complexity)
- **2-Years:** Declines in Sense of Meaningful Work, Task-Content Satisfaction, and Benefits Satisfaction (mirrors techno-invasion & techno-insecurity)
- **3-Years:** Erosion of Overall Duty Satisfaction, with further declines in enjoyment and perceived meaning (signals cumulative toll of techno-uncertainty)

Made with GAMMA

Table 3. First-Stage Results (CHS)

Interval	Statistic	IV1	IV2	IV3	IV4
1 Year	Coef	11.88***	0.318***	0.319***	0.364***
	S.E.	(0.328)	(0.00455)	(0.00445)	(0.00745)
	F	1308.8	4879.3	5123.8	2388.6
	N	2519	2519	2519	2519
	R-squared	0.342	0.660	0.671	0.487
2 Years	Coef	12.72***	0.297***	0.298***	0.343***
	S.E.	(0.369)	(0.00449)	(0.00441)	(0.00740)
	F	1187.7	4376.4	4567.2	2145.1
	N	2307	2307	2307	2307
	R-squared	0.340	0.655	0.665	0.482
3 Years	Coef	12.87***	0.276***	0.277***	0.320***
	S.E.	(0.392)	(0.00438)	(0.00431)	(0.00730)
	F	1077.0	3965.1	4124.1	1915.3
	N	2094	2094	2094	2094
	R-squared	0.340	0.655	0.663	0.478

*, **, *** indicate significance at 10%, 5%, and 1%, respectively. Robust standard errors are in the parentheses.

Made with GAMMA

Table 4. Robotization and Worker Well-Being (CHS)

Interval	Outcome	Statistic	IV2	IV3	N
2 Years	Stress	Coef	0.188**	0.188**	2308
		S.E.	(0.078)	(0.078)	
	Depression	Coef	0.122***	0.121***	2308
		S.E.	(0.042)	(0.042)	
	Monthly Drinking (men)	Coef	0.210***	0.212***	2308
		S.E.	(0.078)	(0.077)	
	Self-Rated Health	Coef	-0.619***	-0.614***	2308
		S.E.	(0.118)	(0.117)	
1 Year	Stress	Coef	0.193***	0.192***	2519
		S.E.	(0.075)	(0.074)	
	Depression	Coef	0.136***	0.135***	2519
		S.E.	(0.040)	(0.040)	
	Monthly Drinking (men)	Coef	0.237***	0.240***	2519
		S.E.	(0.074)	(0.074)	
	Self-Rated Health	Coef	-0.647***	-0.642***	2519
		S.E.	(0.112)	(0.111)	
3 Years	Stress	Coef	0.150*	0.149*	2095
		S.E.	(0.083)	(0.083)	
	Depression	Coef	0.111**	0.110**	2095
		S.E.	(0.045)	(0.045)	
	Monthly Drinking (men)	Coef	0.182**	0.184**	2095
		S.E.	(0.083)	(0.082)	
	Self-Rated Health	Coef	-0.595***	-0.591***	2095
		S.E.	(0.127)	(0.126)	

Note: All estimates are from the IV2 and IV3 specifications (Bartik instruments incorporating major manufacturing economies). Standard errors (SE) are clustered by region.
Significance levels: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 5. Robotization and Job Satisfaction (KLIPS)

Interval	Measure	Statistic	IV2	IV3	N
1 Year	I enjoy doing my current job.	Coef	-5.877**	-5.366**	1928
		S.E.	(2.822)	(2.659)	
2 Years	I find my current job rewarding.	Coef	-2.077*	-2.039*	1672
		S.E.	(1.159)	(1.151)	
	Job content satisfaction	Coef	-2.188*	-2.130*	1672
		S.E.	(1.116)	(1.109)	
	Welfare/benefits satisfaction	Coef	-2.368*	-2.331*	1672
		S.E.	(1.365)	(1.358)	
3 Years	I am satisfied with my current job.	Coef	-2.010**	-2.013**	1491
		S.E.	(0.895)	(0.892)	
	I enjoy doing my current job.	Coef	-1.614*	-1.605*	1491
		S.E.	(0.860)	(0.857)	
	I find my current job rewarding.	Coef	-1.742**	-1.740**	1491
		S.E.	(0.881)	(0.878)	

Note: Table displays only those satisfaction dimensions with statistically significant declines under IV2 and IV3 (Bartik instruments using Singapore, Taiwan, Germany, Japan, and China).
Standard errors clustered by individual.
Significance levels: * p < 0.10; ** p < 0.05; *** p < 0.01.

Table 6. Placebo Test Results: Robotization and Worker Well-Being (CHS)

Interval	Outcome	Statistic	IV2	IV3	N
2 Years	Stress	Coef	-13.033	-12.810	914
		S.E.	(28.221)	(27.921)	
	Depression	Coef	-8.783	-8.971	914
		S.E.	(22.539)	(22.298)	
	Monthly Drinking (men)	Coef	68.734**	67.751**	914
		S.E.	(33.060)	(32.685)	
	Self-Rated Health	Coef	44.175	42.887	914
		S.E.	(51.315)	(50.750)	
1 Year	Stress	Coef	-12.148	-11.777	919
		S.E.	(21.875)	(21.716)	
	Depression	Coef	-10.061	-10.199	919
		S.E.	(17.488)	(17.365)	
	Monthly Drinking (men)	Coef	51.420*	50.491*	919
		S.E.	(25.297)	(25.105)	
	Self-Rated Health	Coef	34.834	33.443	912
		S.E.	(39.852)	(39.542)	
3 Years	Stress	Coef	-17.451	-17.268	912
		S.E.	(38.512)	(38.283)	
	Depression	Coef	-11.508	-11.628	912
		S.E.	(30.885)	(30.701)	
	Monthly Drinking (men)	Coef	94.882**	94.145**	912
		S.E.	(44.738)	(44.455)	
	Self-Rated Health	Coef	62.740	61.730	912
		S.E.	(70.342)	(69.902)	

Robot penetration rate (2012-2019) → **Outcome (2004-2011)**

Table 7. Placebo Test Results: Robotization and Job Satisfaction (KLIPS)

Interval	Measure	Statistic	IV2	IV3	N
1 Year	I enjoy doing my current job.	Coef	0.966	0.963	939
		S.E.	(0.788)	(0.777)	
2 Years	I find my current job rewarding.	Coef	1.640	1.626	969
		S.E.	(1.029)	(1.017)	
	Job content satisfaction	Coef	1.396	1.391	969
		S.E.	(1.004)	(0.992)	
	Welfare/benefits satisfaction	Coef	-0.162	-0.149	969
		S.E.	(1.259)	(1.246)	
3 Years	I am satisfied with my current job.	Coef	-0.411	-0.410	994
		S.E.	(1.660)	(1.646)	
	I enjoy doing my current job.	Coef	-0.092	-0.040	994
		S.E.	(1.587)	(1.574)	
	I find my current job rewarding.	Coef	-0.932	-0.912	994
		S.E.	(1.519)	(1.507)	

Conclusion

1. South Korea, as a highly automated economy, provides crucial insights into the human side of automation.
2. Our study causally links industrial robot adoption to increased stress, depression, alcohol consumption, and decreased self-rated health.
3. Policymakers and firms must accompany robotic investments with robust worker s upports to ensure a human-centered future of work. (**Upskilling initiatives, Job redesign, Mental health resources, Holistic approach**)

Made with GAMMA

THANK YOU



← Contact info.

Made with GAMMA

생성형 AI의 어두운 그림자

이준혁
연세대학교 바른ICT연구소

생성형 인공지능(Generative AI)은 자연어 기반의 직관적인 인터페이스와 즉각적인 피드백을 통해 높은 유용성과 용이성을 제공하며, 보고서 작성, 데이터 분석, 프로그래밍 보조, 창의적 콘텐츠 생성 등 다양한 업무에 활용되고 있다. 이러한 인지된 혜택은 구성원의 업무 효율성과 성과를 향상시키는 긍정적 결과를 가져오기 때문에, 지식노동을 포함한 다양한 직무 영역에서 생성형 AI의 활용이 적극적으로 권장되고 있다. 그러나 생성형 AI의 유용성과 용이성에 대한 인식은 구성원으로 하여금 심리적 의존, 직업 대체에 대한 불안, 역량 개발에 대한 압박감 등 다양한 부정적 감정과 태도를 경험하게 할 수 있다. 이에 본 연구는 생성형 AI의 인지된 혜택 이면에 존재하는 부작용에 대한 경각심을 제고하고자, 감정적·인지적 의존 및 AI 불안을 중심으로 인지된 유용성과 용이성이 직무 스트레스로 이어지는 부정적 메커니즘을 탐색하고자 하였다. 이를 위해 먼저 기술수용모형(TAM 및 UTAUT), I-PACE 모델, 직무요구-자원이론 및 도전–방해 스트레스 프레임워크(CHSF)를 통합하여 개념적 모델을 제시하였으며, 생성형 AI의 유용성 및 용이성 인식 → 감정적·인지적 의존 → AI 불안 → 직무 스트레스로 이어지는 직접효과와 이중 매개효과 가설을 수립하였다. 다음으로 국내 기업 재직자 중 생성형 AI를 업무에 사용하는 사람을 대상으로 설문조사를 실시하였으며, 350부의 유효한 연구 데이터를 확보하였다. 이를 분석한 결과, 인지된 유용성은 인지적 의존을 증가시켰으며, 용이성은 감정적 의존과 인지적 의존을 모두 증가시켰다. 또한, 감정적 의존과 인지적 의존은 AI 불안을 증가시켰으며, AI 불안은 직무 스트레스를 증가시켰다. 마지막으로 감정적·인지적 의존과 AI 불안의 이중매개 효과를 검증한 결과, 인지된 유용성은 인지적 의존과 AI 불안을 매개로 직무 스트레스를 증가시켰으며, 인지된 용이성은 감정적·인지적 의존과 AI 불안을 매개로 직무 스트레스를 증가시켰다. 이상의 연구 결과를 바탕으로, 본 연구는 AI 기술을 조직 내 도입하고, 직무에 통합하는 과정에서 발생할 수 있는 부작용을 최소화할 수 있는 전략적 방안을 제시함으로써 AI 및 인적자원관리 분야에 유용한 이론적·실무적 시사점을 제공한다.

	Memo	

Track
3-4

온라인 리뷰와 eWOM

좌장: 박하은 경희대학교 교수

The Voice of the Player: Sentiment and Topic Shifts in Steam Game Reviews

김엘레나(순천향대학교)
호명휘(순천향대학교)
최재원(순천향대학교)

The Impact of Text and Image Clues on Consumer Trust and Purchase Intention in the eWOM Situation

진우교(순천향대학교)
우여완(순천향대학교)
최재원(순천향대학교)

When Extremes Deceive: The Role of Hedonic and Utilitarian Motivations in Perceiving Manipulation in Online Reviews

이맹름(순천향대학교)
호명휘(순천향대학교)
최재원(순천향대학교)

The Voice of the Player: Sentiment and Topic Shifts in Steam Game Reviews

김엘레나
순천향대학교 경영학과
–
호명휘
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

The rapid growth of the gaming industry underscores the importance of understanding player perceptions in user-generated reviews. This study examines Steam reviews across three stages of the game lifecycle—pre-release, release day, and one-month post-release—using sentiment analysis (VADER) and topic modeling (LDA, DTM). Results show clear shifts in sentiment polarity and thematic focus: anticipation dominates pre-release reviews, while unmet expectations and technical flaws drive negativity post-release. Extreme sentiments cluster around gameplay immersion, monetization, and defects, with a notable rise in extreme negative reviews one month after release. Gameplay duration strongly correlates with sentiment intensity, indicating deeper engagement among invested players. Grounded in Expectancy-Confirmation Theory, UGC, and eWOM, the study advances understanding of digital consumer behavior and offers practical guidance for developers and platforms to anticipate and manage player sentiment throughout the game lifecycle.

KEYWORDS Expectancy-Confirmation Theory, Extreme Sentiment, Game Lifecycle,User-Generated Content (UGC)

The Impact of Text and Image Clues on Consumer Trust and Purchase Intention in the eWOM Situation

진우교
순천향대학교 경영학과
–
유여완
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

The current trend shows that consumers are increasingly looking for relevant information on online social media platforms and looking at eWOM (electric Word of Mouth) of relevant brands or services before choosing goods or services. This study defines "explosive text" as descriptive text that can attract consumer interest. For example, discount beyond the normal range. We refer to descriptive text that captures consumer interest as ‘explosive text '. "Explosive image" usually refers to an image that is exaggerated and attractive, which can attract consumers 'attention, arouse interest and leave a deep impression in a short period of time. Through eWOM related research that affects consumer purchasing behavior, we are considering whether using explosive text/images can more effectively attract consumers' attention. Explosive text/image affects consumers' emotional and cognitive responses. Meanwhile, consumer emotional and cognitive responses have a positive impact on their recommendation intention, consumer satisfaction, user-generated content and digital engagement.

KEYWORDS explosive text/image, eWOM, emotional responses, cognitive responses

When Extremes Deceive: The Role of Hedonic and Utilitarian Motivations in Perceiving Manipulation in Online Reviews

이맹릉
순천향대학교 경영학과
–
호명휘
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

This paper studies the influence mechanism of extreme evaluation in the context of hedonic product (such as virtual goods) purchases. It explores how consumers' hedonic motives regulate their intention to adopt extreme scoring evaluations, and analyzes the specific influence of the content characteristics of the evaluation (such as functional information, timeliness, grammatical correctness, emotional expression, and evaluation length) and the way product information is displayed (such as pictures or videos of product usage scenarios and comparisons with competing products) on this intention. The results of a questionnaire survey of 553 consumers with experience of purchasing virtual goods and structural equation model analysis showed that hedonic motivation significantly and positively affected consumers' adoption of extreme reviews. Among these, rich functional information, high review timeliness, accurate grammar, strong emotional expression, and information such as pictures or videos and comparisons with competing products all further strengthened this positive effect, but excessively long reviews had a negative effect. This study adds to understanding of how online reviews influence consumer decision-making, but also provides practical suggestions for platform operators to focus on the authenticity of reviews and improve the interactivity and entertainment of content.

KEYWORDS Extreme rating comment , Hedonic Motivation , Virtual Goods , Review

Track
3-5

가짜 리뷰 및 허위 정보 탐지

좌장: 박재영 국립부경대학교 교수

A Hybrid Multi-Perspective Approach to Fake Review Detection

사비타(순천향대학교)
가중정(순천향대학교)
최재원(순천향대학교)

Extreme and Deceptive Reviews in Hospitality: Evidence from YouTube Content

부니요드(순천향대학교)
가중정(순천향대학교)
최재원(순천향대학교)

보이스피싱 탐지를 위한 이론적 접근

김재희(순천향대학교)
권성훈(순천향대학교)
김희웅(순천향대학교)

A Hybrid Multi-Perspective Approach to Fake Review Detection

Puri Sabita
Department of Business Administration
Soonchunhyang University

Introduction

Author(s)	Research content
Gupta, D., Bhargava, et.al,2024	The veracity of online reviews, especially those about hotels, is crucial in the digital age. Reviews of hotels have developed into a significant factor in decision-making for tourists. However, the prevalence of false opinions, or "opinion spam," has grown to be a worrying problem in the large sea of hotel evaluations. Travelers may experience financial losses and irritation as a result of false hotel reviews, which can damage the reputation of review sites.
Mukherjee A, Liu B, Glance N,2012	Genuine reviews provide insights into the performance, quality, and overall satisfaction associated with a wide array of offerings, creating a virtual marketplace where transparency and accountability are paramount. However, **the rise of fake reviews poses a significant threat** to the integrity of online platforms and consumer protection.
Bhattacharya R,et.al,2023	Fake reviews can **distort the perceived reputation of businesses, misleading consumers** and potentially leading them to make uninformed decisions.
Barbado R, Araque O, Iglesias C A. 2019	82% of buyers would abandon their purchase decision after reading vast negative reviews about a product, while 87% of consumers would strengthen their shopping intentions based on positive reviews.
et.al.,2023	A UK government report suggests that around 11% to 15% of online reviews may be fake, impacting consumer trust and market competition.
Luca and Zervas,2016	Luca and Zervas (2016) found that on Yelp, the leading restaurant comment site in the US, about 16% of the comments are fake, while on the Amazon platform, the percentage is even higher at 42%.

✓ The harm of fake reviews and their negative impact on user decision-making.

Research Status and Development Trends

Author(s)	Methodology	Research content
Hajek Petr, Barushka Aliak sandr, Munk Michal,2020	Comment text, Characteristics	Hajek et al. proposed a neural network model that integrates the word context and consumer emotion.
Alsubari S Nagi, Deshmukh Sachin N,et.al.,2022	Comment text, Emotional Characteristics	Saleh et al. developed a reviewer detection model that integrated n-grams and sentiment score features.
Yu, S., Ren, J., Li, S., et.al., 2022	Graph structure-driven approach	Graph-based approaches have had remarkable success in the fake review detection process. This success primarily contributed by their ability to establish relationships between features (e.g., reviewers, products/services, reviews, and metadata features).
Xiang, L., You, H., Guo, G., & Li, Q., 2023	Graph structure-driven approach	Introduced a feature fusion method utilizing a heterogeneous graph and performed graph convolution computations to leverage the mutual behavioral insights inherent in the review system.
Qu et al. , 2024	Multimodal fusion applications	Qu et al. addressed the challenges of time sensitivity and the voluminous nature of fake news data by proposing a quantum multimodal fusion-based fake news detection model (QMFND), which outperforms many classical models in terms of detection efficiency.

✓ Existing false comment detection methods can be summarized into three main categories: methods based on text analysis, methods based on graph structural features, and information synthesis methods based on multimodal fusion.

➢ Definition of fake reviews：

✓ Fake Reviews, also known as Opinion Spam, refers to online reviews that do not reflect real user experiences but are disguised as credible evaluations. These reviews may be posted by writers, automated programs, or fake accounts hired by merchants with the primary purpose of misleading potential consumers, manipulating product reputation, or interfering with marketplace fairness (Ott et al., 2011; Mukherjee et al., 2012).

➢ Research gaps:

✓ In summary, existing false comment detection methods show a trend from single to fusion and from static to structural modeling in terms of feature perspectives. However, most of the work still has the following research gaps:

(1) Lack of modeling mechanisms that uniformly fuse textual, structural, and temporal features;

(2) Lack of interpretable detection mechanisms to enhance model transparency and robustness.

➢Research purpose :

This research aims to build a fake review detection framework based on multi-vision fusion that systematically integrates information such as text, structure, and time and improves the model's ability to identify complex fake review behavior.

➢Research questions :

1. How can we build a comprehensive fake review detection feature scheme by simultaneously integrating review text, temporal characteristics, and graph structure information?

2. How to introduce interpretable methods to understand the inspection grounds of the model and enhance the reliability of the system?

Theoretical background

2.1 How to detect fake reviews

Classification	Research content	Author(s)
Text-driven approach	This study used Support Vector Machines (SVMs) combined with text features for fake comment detection and achieved high accuracy on a self-built dataset.	Ott et al. , 2011
Graph structure-driven approach	We propose a large-scale graph convolutional network (GCN)-based anti-spam method for detecting spam advertisements in Idlefish, i.e., a GCN-based anti-spam (GAS) model, which is able to efficiently handle large-scale data.	Li, A., Qin, Z., Liu, R., Yang, Y., & Li, D. , 2019
Supervised learning	This study combines textual and behavioral features to propose a hybrid model that significantly improves detection performance.	Mukherjee et al. , 2013
Unsupervised learning	We focus on the problem of leveraging edge-attributes in social and information graphs for anomaly detection and user behavior modeling purposes. For practitioners, learning about their data in an unsupervised fashion when ground-truth is scarce or unavailable is an important setting, particularly in fraud and anomaly detection use cases.	Shah, N., Beutel, A., Hooi, B., Akoglu, L., Gunnemann, S., Makhija, D., ... & Faloutsos, C. , 2016

✓ Categorized by data type: text-driven methods, graph-structured methods

✓ Categorized by learning mechanism: supervised learning methods, unsupervised learning methods

2.2 Multimodal fusion methods

Early fusion: Fuse different modalities of data in the feature extraction stage to form a unified representation.

Late fusion: Data from each modality is processed separately and finally fused in the decision-making phase.

Author(s)	Research content
Azri, A., Favre, C., Harbi, N., Darmont, J., & Noûs, C. , 2021	We introduce the Multimodal fusiON framework to assess message veracIty in social neTwORks (MONITOR), which exploits all message features (i.e., text, social context, and image features) by supervised machine learning. Extensive experiments are conducted on two multimedia datasets.
Wang et al. , 2018	Wang et al. developed a multimodal Fake News Detection framework, which demonstrated promising results.
Kumar and Taylor , 2023	Kumar and Taylor designed a multimodal fake news detection model that leverages feature engineering to represent the behavior of fake news disseminators.
Wu, Y., Zhan, P., Zhang, Y., Wang, L., & Xu, Z. , 2021	Inspired by the way people read news with image and text, we propose a novel Multimodal Co-Attention Networks (MCAN) to better fuse textual and visual features for fake news detection.
Li, S., Yao, T., Li, S., & Yan, L. , 2022	A fake news detection framework called Semantically Enhanced Multimodal Fusion Network is proposed, which can better capture the common features between events, thus facilitating the detection of fake news.

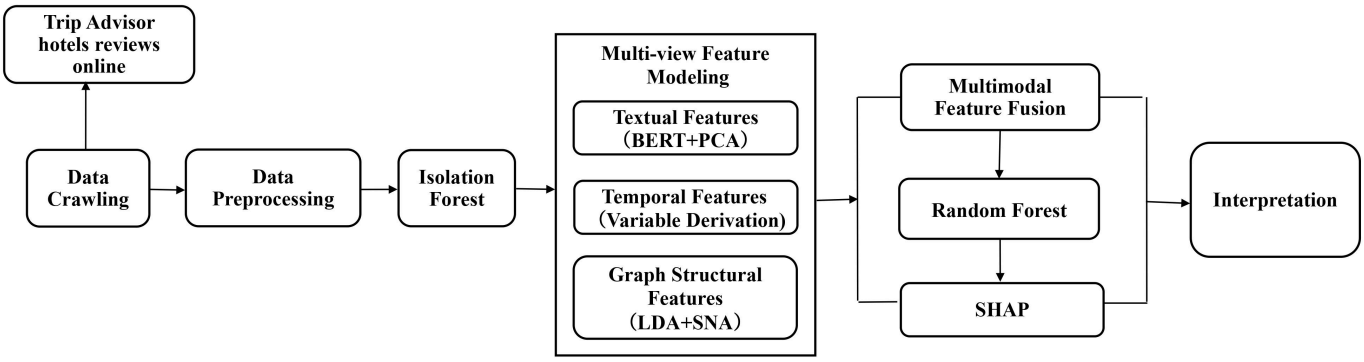
2.3 How to model graph structure and social relationships

Author(s)	Research content
Akoglu et al., 2013	Early work on graph structure modeling focused on constructing bipartite graphs between users and goods, and then applying graph analysis or clustering methods to identify anomalous behavioral patterns .
Zhao, J., Shao, M., Tang, H., Liu, J., Du, L., & Wang, H. , 2023	Zhao et al. proposed a heterogeneous graph consisting of reviews, products, and reviewers for fake reviewers detection, aiming to optimize the mutual information between local and global features. However, they ignored multiple interactions between nodes and mutual correlations among behavioral information, review text, and overall rating.
Li et al., 2019	With the development of deep learning, graph neural networks (GNN) are gradually applied to the task of false comment recognition. In particular, graph convolutional networks (GCNs) and graph attention networks (GATs) are able to automatically learn the embedding representations of nodes and effectively model the complex relationship between the contextual contexts of commenters and the behaviors of their neighbors .
Wang et al., 2020	Some studies have further extended to Heterogeneous GNN (Heterogeneous GNN) by introducing multimodal information such as comment text, timestamps, ratings, etc., to improve the detection performance and generalization ability.

2.4 Interpretable AI methods

Author(s)	Research content
Adadi and Berrada, 2018	Explainable Artificial Intelligence (XAI) has emerged as a pivotal field of research aiming to create transparency in AI model predictions, thereby improving their interpretability. In the context of text classification, interpretability involves understanding why certain texts are classified into specific categories by the model. This understanding aids in uncovering them odel's decision-making process and helps in trust building among its users.
Mohammed & George,2023	SHAP (Shapley Additive ExPlanations) is an Explainable AI (XAI) approach for interpreting machine learning model output .
Weitz et al. ， 2021	SHAP has been utilized in several studies to provide transparency in text classification. For instance, Weitz et al. explored the potential of virtual agents in XAI designs on the perceived trust of end-users.
Lundberg and Lee 2017 ; Lundberg et al. , 2018	We use the Shapely Additive Explanations (SHAP) method as it has a solid theoretical foundation that provides local accuracy and consistency in real-world scenarios and it is an agnostic interpretability technique that can be used to explain any deep learning model.

Research Procedure



Data Collection

➢Data collection: Trip Advisor reviews for hotel reviews (collection time: 2020-2025)

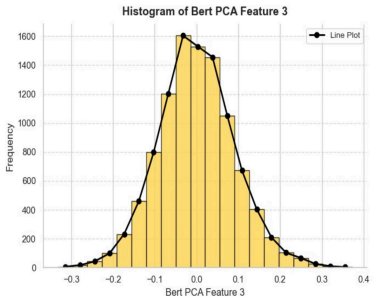
➢A total of 59 hotels were collected, with a total of 101,020 reviews.

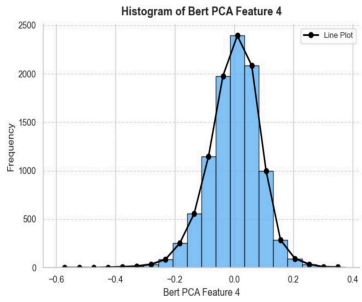
➢Data Pre-processing

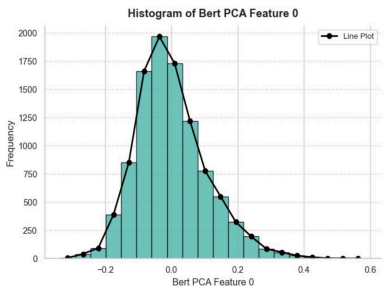
1. Change text to lowercase ;
2. Replace duplicate punctuation: Remove all non-alphabetic characters such as symbols and spaces ;
3. tokenization: splits text into individual words ;
4. Remove tokenizers that contain numbers or non-alphabetic characters ;
5. Removing stop words: removes commonly used words such as “the”, “is”, “in”, etc.

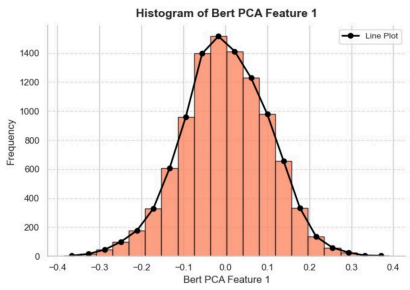
Textual Features(BERT + PCA)

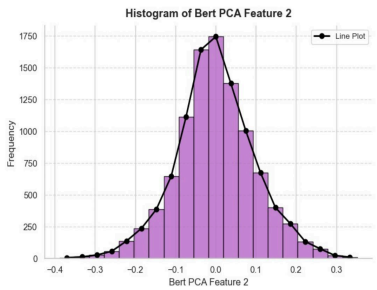
➢BERT Feature Extraction
➢PCA Dimensionality Reduction
➢Rate-derived emotional features











Temporal Features(Variable Derivation)

- ➢1. Standardize time fields (harmonize format to facilitate extraction of time attributes)
- ➢2. Extraction of comment month (captures monthly distribution of comment activity)
- ➢3. Time interval from the first comment (characterizes the "time distance" of comments; identifies concentrated outbursts or delayed behavior of comments)
- ➢4. Normalized time position (normalizes comment time to a 0-1 interval; useful for user modeling)

Graph Structural Features(LDA + SNA)

1. LDA Topic Modeling
2. Creating Topic-Based Review Plots

- ✓ Each review is represented by its topic distribution vector.
- ✓ Compute the topic similarity between reviews to construct an undirected map between reviews.



Table1:Organizing LDA topics

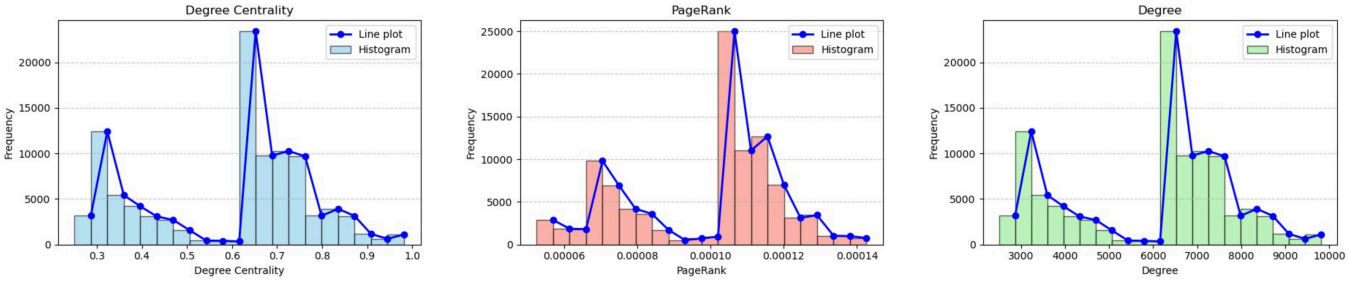
	Quality of service and experience		Ambience and Dining Experience		Vacation and leisure facilities	
	Word	Weight	Word	Weight	Word	Weight
1	hotel	0.0400	nice	0.0367	pool	0.0157
2	staff	0.0328	staff	0.0308	hotel	0.0147
3	stay	0.0246	friendly	0.0290	resort	0.0113
4	hoi	0.0196	clean	0.0261	staff	0.0096
5	breakfast	0.0155	breakfast	0.0228	stay	0.0077
6	pool	0.0152	pool	0.0219	day	0.0072
7	town	0.0141	hotel	0.0193	service	0.0067
8	friendly	0.0138	time	0.0191	view	0.0058
9	helpful	0.0128	delicious	0.0168	bar	0.0053
10	service	0.0121	view	0.0156	breakfast	0.0052

3. Graph Structure Feature Extraction

✓Degree Centrality: Indicates the degree of connection between nodes

✓PageRank: Measures the importance of a node in the overall picture

✓Degree: The number of edges a node has



An RF-based method for identifying fake reviews: Fusion graph structure, temporal and textual features.

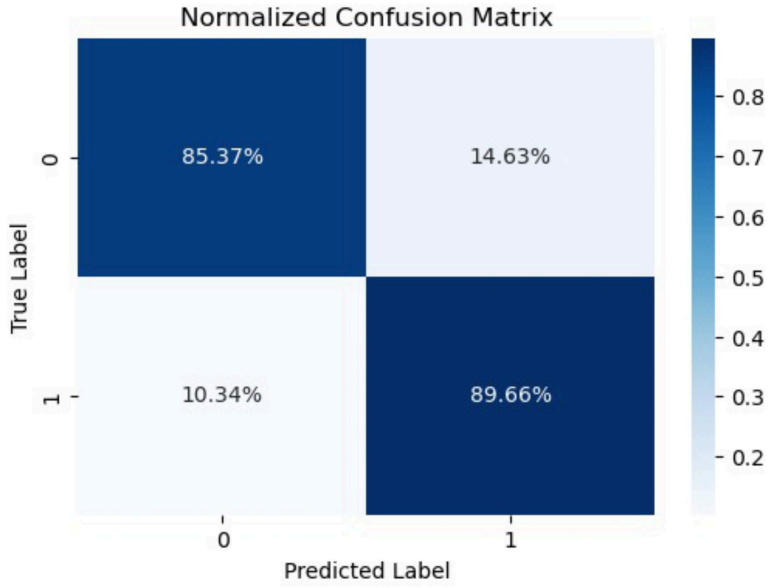
✓ In the model evaluation, the training set was matched with the test set, and the accuracy of the test set was 87.54% and the F1-Score was 87.93%, showing good generalization ability.

✓ Among them, the recovery rate of fake reviews recall 89.66%, indicating that the model has strong discrimination ability and can effectively detect fake reviews.

	Accuracy	Precision	Recall	F1-Score
Training	0.8750	0.8761	0.8721	0.8741
Text	0.8754	0.8627	0.8966	0.8793

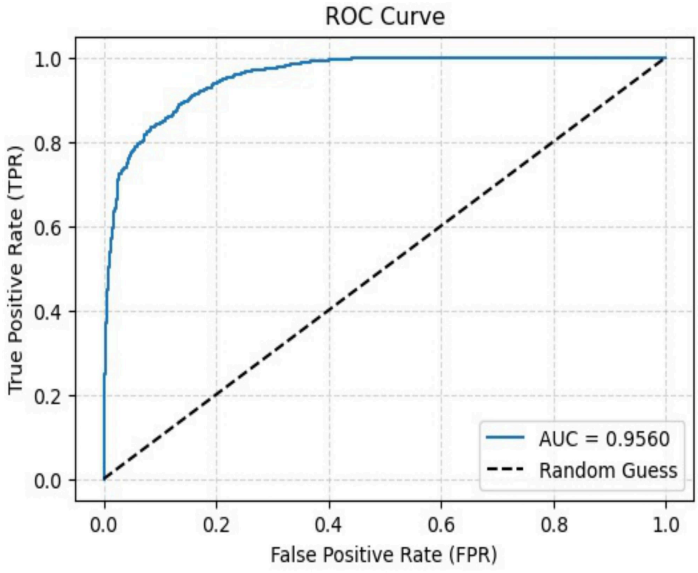
Normalized Confusion Matrix

- ✓ Both false positive (FP) and false negative (FN) detections are relatively low, with particularly good identification of false reviews (TP is high at 89.66%), indicating that the model has a strong ability to identify fake reviews.



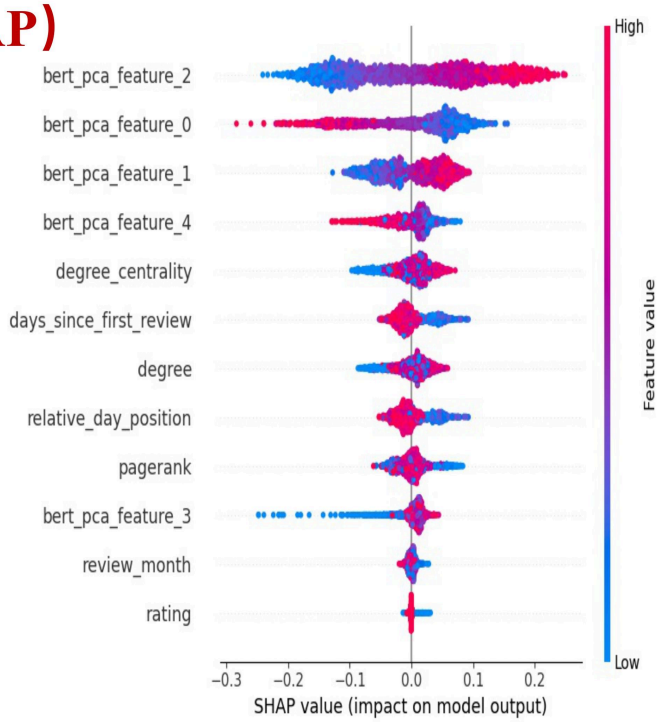
ROC Curve

- ✓ AUC = 0.956 indicates that the model has strong discriminative power.
- ✓ Real and fake reviews were randomly selected and the model correctly distinguished between them 95.6% of the time.
- ✓ The ROC curve is skewed toward the upper left corner throughout, indicating that the model is stable for most threshold settings.

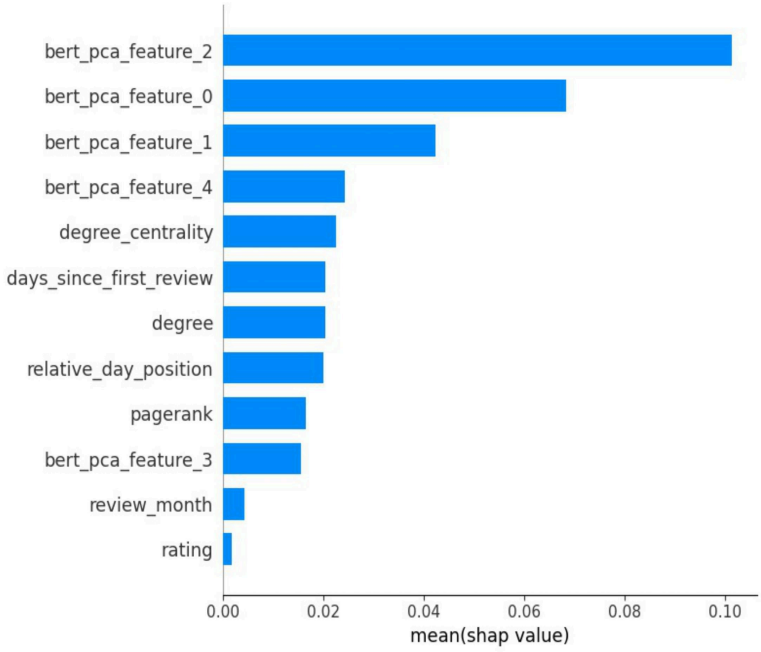


Interpretable analysis（SHAP）

- ✓ Bert-pca-feature-2 has the largest impact on the model due to the most widely distributed SHAP values, with high values generally favoring model predictions toward the positive class (possibly fake reviews).
- ✓ Several bert-pca-feature attributes have a significant impact.
- ✓ Related attributes of the graph structure, such as degree-centrality, degree, and days-since-first-review, also have a larger impact, indicating that the user review network structure also plays a role in the model.
- ✓ The small impact of rating and review-month suggests that rating and review time have a limited predictive contribution to the model.



- ➢ The average contribution of each feature to the model.
- ➢ Used to rank feature importance and quickly identify dominant features.



Conclusion

- The multimodal fusion falke comment detection method proposed in this paper combines textual, graph structural and temporal features based on random forest model with good results.
- The model performs well on the test set with 87.54% accuracy and 87.93% F1-Score, with strong generalization ability and robustness.
- The fake comment recall rate is 89.66%, effectively identifying most of the fake comments with a low fake alarm rate. The AUC value is 0.9566, indicating that the model is stable and accurate under various thresholds.
- SHAP analysis shows that BERT dimensionality reduction text features are the main decision basis, graph structure features contribute significantly, and time and scoring have less impact. The model interpretability enhances the transparency and credibility of the system and provides direction for subsequent optimization.

Thank You !!!

Extreme and Deceptive Reviews in Hospitality: Evidence from YouTube Content

부니요드
순천향대학교 경영학과
–
가중정
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

This study aims to identify and filter extreme and fake hotel reviews on the YouTube platform using machine learning and natural language processing (NLP) techniques. As social media increasingly influences consumer decisions, user-generated reviews have become a key factor in online marketplace credibility. However, review manipulation and fake reviews threaten platform trust. This research contributes to enhancing review authenticity and supporting rational consumer decisions.

KEYWORDS Fake reviews, YouTube, Word2vec, Sentiment classification,Latent Dirichlet Allocation

	Memo	

2025 한국지식경영학회 추계학술대회

A Theory-Guided Approach to Voice Phishing Detection

보이스피싱 탐지를 위한 이론적 접근

2025년 11월 13일 (목)

김재희, 권성훈, 김희웅

연세대학교 정보대학원

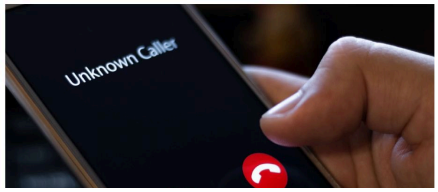
■ Voice Phishing as Crucial Problems

– The Growing Severity of Voice Phishing

- Voice phishing has become a critical global issue, resulting in financial and psychological damages
- Global voice phishing incidents surged by 550% in 2022 (ABHandshake, 2024)
- Approximately 300,000 phishing victims in the U.S. suffered losses amounting to $52 million (Forbes, 2023)
- 'Deep Voice', an emerging AI-generated tool, poses a growing threat (New York Post, 2024)

– Psychological and Societal Impact of Voice Phishing

- Victims often experience guilt, depression, and anxiety disorders due to the financial damage (Straits Times, 2023)
- Phishing can erode trust between individuals, financial institutions and government agencies, ultimately weakening social capital (Anderson et al., 2014)



■ Limitations of Prior Approaches

– Challenges in Existing Phishing Detection Models

- Need for automated analysis to address evolving phishing methods (The guardian, 2024)
- Reliance on call number history for initial detection approaches (Jiang et al., 2012; Murynets et al., 2014)
  - Challenges in detection due to infinite number generation, internet phone bypassing, and advancements in phishing techniques
- Shift toward AI-based detection models (Peng & Lin, 2019; Zhao et al., 2018)
  - E.g., Mobile Carriers' Call Apps: using phone number activity patterns for phishing detection

– Limitations of Machine Learning Approaches

(1) Vulnerability to Evolving Phishing Techniques

- Continuous changes and advancements of phishing tactics make detection ineffective, requiring frequent model retraining (Peng & Lin, 2018; Zhao et al., 2018)

(2) Lack of Theoretical Foundation

■ Limitations of prior approaches

– Challenges in Existing Phishing Detection Models

- Need for automated analysis to address evolving phishing methods (The guardian, 2024; DongA Science, 2023)
- Reliance on call number history for initial detection approaches
  - Challenges in detection due to infinite number generation, internet phone bypassing, and advancements in phishing techniques
- Shift toward AI-based detection models (Peng & Lin, 2019; Zhao et al., 2018)
  - E.g., Mobile Carriers' Call Apps: using phone number activity patterns for phishing detection

To address these challenges, we aim to

Develop a voice phishing detection framework that incorporates *theoretical foundations* to capture the comprehensive context

– The Growing Severity of Voice Phishing

(1) Vulnerability to Evolving Phishing Techniques

- Continuous changes and advancements of phishing tactics make detection ineffective, requiring frequent model retraining (Peng & Lin, 2018; Zhao et al., 2018)

(2) Lack of Theoretical Foundation

■ Related Works

- Existing researchs focus on **data-driven methods** for performance but lacks contextual understanding of situations
- Applying the concept of **'Threat'**, we emphasize **phisher's explicit and implicit messages** that trigger protective behaviors and enhance adaptability across diverse phishing scenarios

<Table 1> Review of Prior Literature on Voice Phishing Detection

Research Approach	Author	Methods			Key Features
		ML	DL	Other	
Call Pattern Based Approach	Jiang et al. (2012)			✓	Voice call history (Markov Clustering)
	Murynets et al. (2014)	✓			Voice call history (Network traffic)
Keyword Based Approach	Peng & Lin (2018)			✓	Text frequency (Fraud community)
	Zhao et al. (2018)	✓		✓	Text frequency (Social phishing keyword)
	Lee & Park (2021)	✓			Text frequency (Real voice phishing keyword)
Voice	Rahman & Bandung (2022)	✓			Voice signal (MFCC)
	Fernandes & Ullah (2021)		✓		Voice signal (MFCC)

■ The concept of Threat

– Theoretical Foundation: The concept of Threat

- 'External stimulus variables' that exist regardless of an individual's awareness (Witte, 1992)
- Distinct from natural dangers or business risks; it elicits defensive responses from victims (Loch et al., 1992; Battistelli & Galatino, 2018)
- Fear appeal messages significantly amplify threat perception: emphasizing both severity and self-vulnerability (Witte, 1992)

– Prior Studies Focusing on **'Perceived Threat'**

- Maily examined users' motivations triggered by perceived external threats to reduce risks from cyber attacks (Martens et al., 2019; Tsai et al., 2016)
- Implicit Threats: Highilgiht the importance of recognizing externally existing threats that are not immediately visible for proactive and preventive measures

■ Interpretability with XAI

– Theory-guided Design: Design Principles for Interpretability

- Integrate the concept of 'threat' to define explicit vs. implicit threat cues for interpretable feature design
- Evaluate interpretability using SHAP-based explanation

– Advancement of XAI-based Methodology

- Develop graph-based relational explanations (e.g., GNN attention for threat–response interaction visualization)
  - Capture interactions between explicit and implicit features during phishing situations (Guo et al., 2024)
- Benchmark model-agnostic methods (LIME, SHAP, IG) for robustness and consistency in explanations

– LLM-aided Inference Enhancement

- Combine XAI and LLM for an inference stage that tests real-world decision support
- Produce LLM-based textual explanations and validate with experts (e.g., investigators, etc.)
- Build a multi-layered explanatory framework integrating diverse threat cues

Jaehui Kim (Ph.D Student)
Graduate School of Information, Yonsei University, Seoul, South Korea
Digital Services Lab

	Memo	

	Memo	

	Memo	

	Memo	

2025 공동 학술대회

Track 4

Track 4-1	Track 4-2	Track 4-3	Track 4-4	Track 4-5
대규모언어모델과 ESG 경영 좌장: 이규민 동국대학교 교수	온라인 플랫폼의 위험관리 전략 좌장: 이중엽 소프트웨어정책연구소 책임연구원	전기차관련 인식 및 구매 의사결정 좌장: 고준 전남대학교 교수	브랜드 스토리와 AI 인플루언서 좌장: 윤상혁 동국대학교 교수	디지털 업무환경과 서비스 좌장: 박상철 대구대학교 교수
메타인지 기반 접근을 통한 대화형 인공지능의 멀티턴 응답 품질 개선 박준형(경희대학교), 김영혜(경희대학교), 박선우(경희대학교), 이청용(한성대학교), 양성병(경희대학교)	소셜미디어 플랫폼의 악성 콘텐츠 대응 전략 임희주(서원대학교)	소비자 불만이 전기차 위험 인식과 회피 의도에 미치는 영향: HBM과 TTAT 관점에서 왕문준(순천향대학교), 최재원(순천향대학교)	스토리의 힘: 디지털 시대 브랜드 스토리텔링이 소비자–브랜드 간 심리적 유대를 형성하는 메커니즘 유몽(순천향대학교), 최재원(순천향대학교)	생성형 AI 시대 개발자의 과의존과 중독 안재영(연세대학교), 채송화(KT DS), 임혜정(연세대학교), 윤혜정(이화여자대학교)
멀티모달 대규모 언어 모델의 시각 이해 개선을 위한 입력 모달리티 조합에 대한 실증 연구 김도원(국민대학교), 조윤호(국민대학교), 문현실(국민대학교)	AI 모델의 프라이버시-신용 위험 예측 정확도 상충 관계 분석을 통한 소비자 신용 위험 관리 양은경(연세대학교)	Exploring the Expectation–Mismatch in Electric Vehicle Adoption: Insights from Negative Consumer Reviews on Reddit 정자강(순천향대학교), 왕문준(순천향대학교), 최재원(순천향대학교)	Exploring User Acceptance of AI Influencers: A Big Data Analysis of Trust, Engagement, and Brand Influence through Instagram 시함(순천향대학교), 최재원(순천향대학교)	디지털 인증 맥락에서 본 그림자노동의 구조적 전환: 행위자-네트워크 이론 관점의 분석 박상철(대구대학교), 고준(전남대학교)
디지털 기업 창업자와 투자자 간 ESG 경영요소 도입 우선순위 인식 비교에 관한 연구: 혼합 방법론 사용 박세민(전남대학교), 고준(전남대학교)	From YouTube to Reddit: Cross-Platform Topic Propagation in World of Warcraft 르키아(순천향대학교), 유여완(순천향대학교), 최재원(순천향대학교)	플랫폼 기반 타이어 구매 의도의 결정요인과 쇼루밍 성향의 조절효과 김상준(경희대학교), 양성병(경희대학교)	Automatic Detection Thumbnail–Title Congruence in YouTube Branded Videos using Multimodal Machine Learning: Implications for View-Based Engagement 르키아(순천향대학교), 최재원(순천향대학교)	호텔 종사원의 사회적 평판과 자기조절효능감이 서비스 대응성에 미치는 영향: 사회인지이론관점 권나현(경희대학교), 정남호(경희대학교)
	Behind the Stars: A Text Mining Study of Negative Reviews and Seller Strategies 김엘레나(순천향대학교), 최재원(순천향대학교)			

	Memo	

Track
4-1

대규모언어모델과 ESG 경영

좌장: 이규민 동국대학교 교수

메타인지 기반 접근을 통한 대화형 인공지능의 멀티턴 응답 품질 개선

박준형(경희대학교)
김영혜(경희대학교)
박선우(경희대학교)
이청용(한성대학교)
양성병(경희대학교)

멀티모달 대규모 언어 모델의 시각 이해 개선을 위한 입력 모달리티 조합에 대한 실증 연구

김도원(국민대학교)
조윤호(국민대학교)
문현실(국민대학교)

디지털 기업 창업자와 투자자 간 ESG 경영요소 도입 우선순위 인식 비교에 관한 연구: 혼합 방법론 사용

박세민(전남대학교)
고준(전남대학교)

메타인지 기반 접근을 통한 대화형 인공지능의 멀티턴 응답 품질 개선

Metacognition-Based Approach to Enhancing Multi-turn Response Quality in Conversational Artificial Intelligence

2025. 11. 13

경희대학교 일반대학원 빅데이터응용학과 석사과정 박준형

#0. 목차 (Context)

:메타인지 기반 접근을 통한 대화형 인공지능의 멀티턴 응답 품질 개선 연구

1. 서론 (Introduction)

2. 이론적 배경 (Theoretical Background)

3. 연구 방법 (Research Methods)

4. 예상되는 분석 결과 (Expected Analysis)

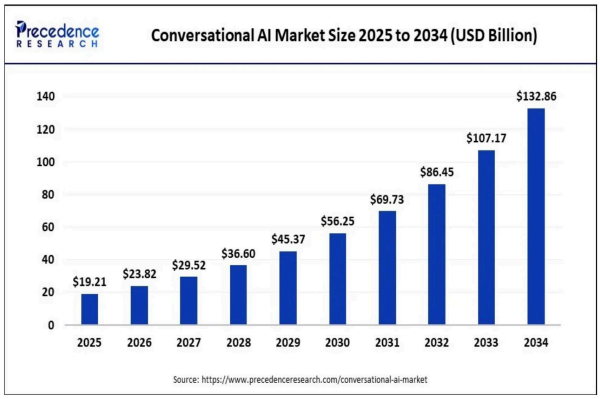
5. 예상되는 결론 (Expected Conclusion)

#1. 서론 (Introduction)

: 1-1 연구배경

대화형 AI 등장 배경

- 최근 GPT, LLaMA 등 대규모 언어 모델(Large Language Model, LLM)의 발전과 함께 생성형 AI(Genreative AI)가 급격하게 확산되고 있음(Cusumano, 2023).
- 초기에는 사용자의 지시사항을 이해하고 텍스트, 이미지 등 다양한 유형의 결과물을 고품질로 생성하는 능력에 주목함 (Mariani & Dwivedi, 2024).
- 그러나 실제 환경에서의 문제는 **요구사항이 불완전한 경우가 대부분**이며, 이를 해결하기 위해서는 반복적인 상호작용이 필요함(Zaib et al., 2022).
- 실제로 **인간은 복잡한 문제를 해결**할 때, **질문과 응답을 반복**하며 사고를 정교화하는 경향이 있음(Chi & Wylie, 2014).
- 이에 따라 대화 형식의 응답을 생성하고 맥락을 유지하며 사용자와 상호작용하는 **대화형 AI가 주목받고 있음**(Gao et al., 2019; Ni et al., 2023).



<그림 1> 대화형 AI 시장 전망(Precedence Research, 2025)

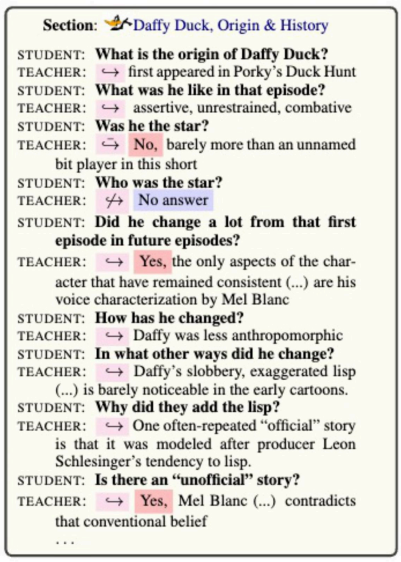
#1. 서론 (Introduction)

: 1-1 연구배경

멀티턴 대화형 AI의 특징

- 대화형 AI는 단순한 응답 생성에 그치지 않고, **질의, 응답, 피드백의 순환**을 통해 문제 해결 과정을 정교화함(Venter et al., 2025; Fu et al., 2022).
- 특히, 생산성과 정확도를 개선하고 비용을 절감시키는 효과로 인해, 다양한 응용 환경에서 표준 인터페이스로 빠르게 채택되고 있음(Zhao et al., 2023; Brynjolfsson & Raymond, 2025; Cusumano, 2023).
- 이러한 대화형 AI를 효과적으로 운용하기 위해서는 **멀티턴(multi-turn) 대화 구조**가 필수적임(Reimann et al., 2024; Yang et al., 2022).
- 단일턴(single-turn)은 개별 질문에 대한 단발적 응답에 한정되므로 복잡한 과업 해결에 한계가 있음(Li et al., 2025).
- 반면, 멀티턴은 대화를 통해 **맥락 누적, 오류 교정, 단계별 정교화**를 가능하게 하여 대화형 AI의 신뢰성과 활용성을 강화함(Zhang et al., 2025).

➢ ***그러나 멀티턴은 장점과 동시에 구조적 취약성도 내포하고 있음.***



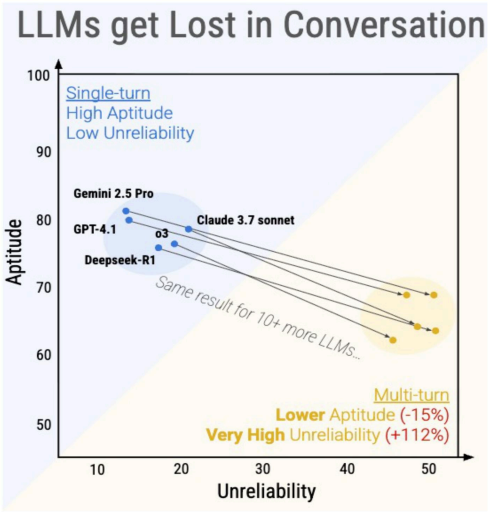
<그림 2> 멀티턴 대화 예시(Choi et al., 2018)

#1. 서론 (Introduction)

: 1-1 연구배경

멀티턴 대화형 AI의 한계점

- 실제로 대화형 AI는 **환각(hallucination) 문제**로 인해 **멀티턴 환경의 불안정성**이 지속적으로 지적되고 있음(Farquhar et al., 2024; Huang et al., 2025).
- 대표적으로, **초기 오류**가 연쇄적으로 확산되는 오류 전파(Error Propagation)는 잘못된 출력을 기반으로 새로운 환각을 생성하는 자기증폭적 오류(self-reinforcing error)를 유발할 수 있음(Kwan et al., 2024; Chen et al., 2025).
- 또한, 모델은 스스로 이전 응답을 점검하거나 불일치를 감지하지 못해, 발화 간 상충되는 정보가 누적될 경우 **맥락 불일치 (Contextual Inconsistency)**가 나타날 수 있음(Xue et al., 2023).
- 이러한 문제로 인해 질문의 본래 의도가 왜곡되고 대화 전체의 신뢰성이 저하되며, 이는 안전 문제, 법적 분쟁, 경제적 손실로 이어질 수 있음(NIST, 2024).



<그림 3> 멀티턴 환경에서의 불확실성 양상(Laban et al., 2025)

#1. 서론 (Introduction)

: 1-1 연구배경

멀티턴 환경 문제 해결 연구

- 기존 연구들은 환각 문제를 완화하기 위해 다양한 접근법을 제안해왔으며, **개입 시점**에 따르면 크게 세 가지로 분류할 수 있음.
- 첫째, **사후(post-hoc) 접근법**은 이미 생성된 출력을 평가·수정하는 방식으로, 대표적으로 불확실성 정량화와 자기 일관성 검증이 있으며, 독립적 샘플링을 통해 신뢰성을 보강하고자 함(Huang et al., 2025; Agrawal et al., 2023).
- 둘째, **사전(pre-input) 접근법**은 모호하거나 복잡한 질의를 재작성(query rewriting)하여 환각 발생 가능성을 사전에 줄이는 방법으로, 질의의 명확화를 통해 환각 발생을 예방할 수 있음(Kim et al., 2024; Mo et al., 2025).
- 셋째, **중간(in-process) 접근법**은 생성 과정 중 reasoning trace 점검이나 chain-of-thought 기반 평가를 수행하여, 출력이 완성되기 전에 환각을 억제하는 방식임(Pozzi et al., 2025).

#1. 서론 (Introduction)

: 1-1 연구배경

기존 연구의 한계점

- 그러나 이러한 접근법들은 공통적으로 **단일 단계에서만의 성능 개선**이라는 한계를 가짐.
- 사후 검증은 누적 맥락을 반영하지 못하며, 질의 재작성은 무차별 적용 시 원본 의도를 왜곡하거나 환각을 오히려 증가시킬 수 있음(Roy et al., 2024; Zhang et al., 2024).
- 또한, 멀티턴 환경에서는 질의의 위험도와 복잡성을 사전에 평가하여 적절한 생성 전략을 선택해야 하나, 기존 기법들은 이러한 예방적 관점을 충분히 반영하지 못함(Su et al., 2025).
- 마지막으로, LLM의 자기회귀적 특성상 초기 오류가 후속 토큰 전반에 연쇄적으로 영향을 미치기에 **생성 전 과정에 걸친 실시간 모니터링이 필요함**(Laban et al., 2025; Sun et al., 2023).
- 그러나 **기존 연구는 출력 이후 단계에 집중**하며, 여전히 멀티턴 대화의 구조적 불안정성으로 인한 환각을 해결하는데 한계가 존재함(Manakul et al., 2023).

#1. 서론 (Introduction)

: 1-2 연구의 필요성 및 목적

연구의 필요성

- **멀티턴 대화는 오류 전파·맥락 불일치로 환각이 누적되어 단일턴 대비 불안정성이 커지므로 이를 선제적으로 감지·완화할 체계가 필요함.**
- **기존 기법들은 단일 질의응답에 최적화되고 전 주기적 모니터링을 통한 예방적 제어를 진행한 연구가 부족함.**
- **LLM의 자기회귀 특성상 초기 오류가 후속 토큰에 연쇄 전이되므로, 생성 전 과정을 아우르는 감시·개입이 요구됨**

➢ ***멀티턴에서의 환각문제는 계속 언급되고 있고 연구되어 왔으나 생성 전 과정을 모니터링하는 연구는 부재함.***

연구 목적

- **Nelson & Narens(1990)의 메타인지를 통한 접근법으로 인지과정의 단계별 지표를 정량화해 '메타 점수'를 산출하고, 산출된 점수를 기반으로 질의 재작성 여부를 결정함.**
- **객체 수준과 메타 수준 루프를 설계하여 상향 모니터링을 통해 질의 재작성 여부를 실시하는 제어 모듈을 구현하여, 멀티턴 응답의 품질을 향상하고자 함.**
- **제안한 프레임워크를 공개된 벤치마크 데이터셋인 MTRAG로 답변 품질을 평가하여 본 연구의 프레임워크의 효과성을 입증하고자 함.**

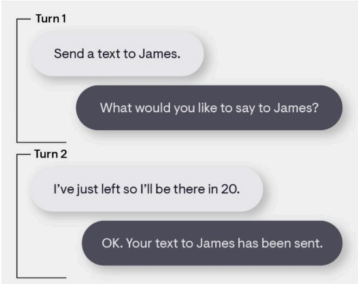
➢ ***본 연구에서는 Nelson & Narens(1990)의 메타인지에서 착안해서 인지과정 전체를 답변 생성 과정 전체에 대응시켜 쿼리 재작성이 필요한 쿼리만을 선별하도록 하는 프레임워크를 제안하고자 함.***

#2. 이론적 배경 (Theoretical Background)

: 2-1 대화형 AI

2.1 대화형 AI

- 생성형 AI는 콘텐츠 창작 기술로 발전했으나 상호작용 능력이 제한되었던 반면, ChatGPT의 등장으로 대화형 AI가 발전하면서 인간과 자연스러운 대화가 가능해짐(Dwivedi et al., 2023).
- 대화형 AI는 생성형 AI에서 더 나아가, 멀티턴 상호작용으로 의도를 명확히 묻고 맥락을 축적하여 **실제 과업을 정확하게 완수**하게 만들며, 이런 이점 때문에 최근 산업과 연구에서 대화형 AI가 표준으로 부상함(Zamani et al., 2023; Mo et al., 2024).
- 대화형 AI는 대화 상태를 추적하여, 그 상태에 따라 대화 정책이 다음 행동을 선택해 답변을 생성하며, 이는 **인간의 메타 인지를 통한 진행 상황을 모니터링하고 그에 맞춰 행동을 결정하는 방식**과 유사함(Jacqmin et al., 2022; Ackerman & Thompson, 2017).
- 싱글턴은 한 번의 질의에 대한 응답에 머물러 복잡한 과업에서 맥락과 의도를 반영하기 어려움(Aliannejadi et al., 2019).
- 하지만, 멀티 턴에서는 대화 상태를 반영하여 행동을 결정하므로, 맥락 누적·모호성 해소·오류 교정 등 **사용자와의 대화를 통해 복잡한 작업을 수행할 수 있게 함**(Mo et al., 2024).



<그림 4> 멀티턴 대화 예시(PolyAI, 2024)

#2. 이론적 배경 (Theoretical Background)

: 2-2 멀티턴 대화형 AI의 한계

2.2 멀티턴 대화형 AI의 한계

- 이러한 멀티턴에서의 환경의 우수성에도 불구하고, **대화형 AI는 구조적 한계를 지니며 이 한계는 멀티턴 환경에서 심화됨**.
- 오류 전파와 자기증폭적 환각: 초기 턴의 작은 오류가 후속 턴 입력으로 재주입되며 전반적인 성능이 단일턴 대비 크게 하락하는 문제가 발생할 수 있음(Laban et al., 2025).
- 장기 맥락 유지 실패와 맥락 불일치: 멀티턴 대화에서 대화형 AI는 중간의 턴을 반영하지 못하고 마지막 턴에 과도 적응하는 경향을 보이며, 이전 잘못된 시도를 반영하여 답변의 일관성이 깨져 정확도가 저하되는 현상이 발생하고 있음(Kwan et al., 2024; Laban et al., 2025).
- 불확실성 오평가: 대화형 AI의 과신과 불안정한 불확실성 추정이 멀티턴 상호작용에서 정책 결정을 왜곡함(Li et al., 2025).
- 또한, 멀티턴에서 의도·주제가 변하며 검색·생성이 엇갈리는 문제가 빈발하는 문제가 지적되고 있음(Cheng et al., 2024).

#2. 이론적 배경 (Theoretical Background)

: 2-3 멀티턴 대화 환경

2.3 멀티턴 대화 환경

<표 1> 멀티턴 대화 관련 선행 연구

단계	연구 방법	한계점	저자(년도)
생성전	페르소나 기반 사용자 모델링으로 어떤 정보를 언제 어떻게 반영할지 설계해 멀티턴의 개인화·유창성을 개선	페르소나 주입에 집중되어 턴 간 사실성/오류 모니터링의 부재	Kwon et al. (2023)
	문백 기반 질의 재작성으로 지시어/대용어를 해서하고 선호 정렬 학습으로 사용자 의도에 맞춘 입력을 재구성	입력(질의)만 다루어 출력의 사실성·정합성 검증이 부재	Sun et al. (2022)
	명확한 질문을 통해 모호성을 해소하고 강화학습 기반으로 언제/어떻게 되물을지 학습	모호성 해결에 치중하여 사실 오류·사실성 검증 부재	Chen et al. (2020)
생성중	계층형 문맥 모델링과 변화 추적 모듈로 멀티턴 일관성 강화	구조적 일관성을 향상되나 실시간 모니터링과 그에 대한 교정이 부재	Yang et al. (2022)
	지식 그래프 근거화와 디코딩 중 사실 주의, 최종 환각도 기반 재랭킹으로 사실성 제고	지식 그래프에 의존하고 엔티티 연결시 취약하며 사후 교정 부분이 미흡	Ji et al. (2022)
	외부 모메리를 통합하여 장기 개인화 맥락을 유지하고 검색 전략을 강화학습으로 조정	메모리·개인화 중심으로 사실 검증이 부재	Tan et al. (2023)
생성후	의미 엔트로피로 생성 결과의 환각 가능성을 사후 탐지	탐지만 하고 교정을 하지는 않음	Farquhar et al. (2024)
	다중 샘플 자기일관성 점검으로 사실성 추정	샘플링 비용이 크며, 초기 오답 노출 후 탐지, 모델이 일관되게 틀리면 탐지가 어려움	Manakul et al. (2023)
	자기피드백 기반 반복 보정으로 초기 답을 스스로 비평·수정해 품질 향상	대화 전체 관점 부족, 반복 생성으로 지연 증가	Madaan et al. (2023)

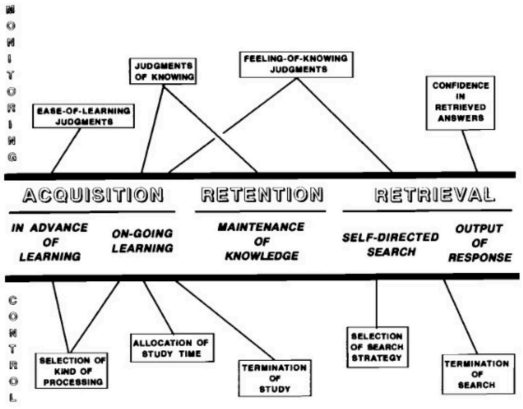
➢ ***위 방법들은 멀티턴의 개별 고질 문제를 단일 단계에서 부분적으로 다루지만 답변 생성 전 과정을 모니터링하여 답변이나 질의를 수정하는 제어를 수행하는 '전주기적 제어' 연구는 부재함.***

#2. 이론적 배경 (Theoretical Background)

: 2-3 인간의 메타인지

2.3 인간의 메타인지

- 메타인지는 '**자신의 인지 과정을 인식·평가**하고 그 결과에 따라 **전략을 조절하는 능력**'으로 정의됨(Flavell, 1979).
- 메타인지는 **문제 해결**, 교육, 의료 진단 점검 등의 인지 분야에서 활용됨(Rivas et al., 2022).
- 특히, Nelson & Narens(1990)은 인간 인지를 **객체 수준(object-level)과 메타 수준(meta-level)**으로 구분함(Nelson & Narens, 1990).
  - **객체 수준(object-level):** 실제로 인지 처리가 이루어지며 질문에 답을 하는 등의 실제 행동을 수행
  - **메타 수준(meta-level):** 자신의 인지 과정을 관찰하고 판단하는 수준
  - **상향 보고(monitoring):** 객체 수준의 상태가 메타수준으로 전달됨
  - **하향 제어(control):** 메타 수준의 판단이 객체 수준의 행동을 조절함
- **학습 전 예상(EOL, Ease of Learning), 학습 중 판단(JOL, Judgement of Learning), 회상 후 확신도(FOK, Feeling of Knowing)**가 연속적으로 이루어지며, 각 판단이 자기 점검 전략에 실질적 영향을 미침(Dunlosky & Metcalfe, 2008).



<그림 5> 메타인지 기억 체계에서 주요 단계(Nelson & Narens, 1990)

#2. 이론적 배경 (Theoretical Background)

: 2-3 Nelson & Narens(1990) 단계별 정의와 본 연구로의 대응

2.3 Nelson & Narens(1990) 단계별 정의와 본 연구로의 대응

- 메타인지는 **'학습'**만이 아니라 **'문제 해결 과제'**에서도 그대로 적용할 수 있음(Zepada & Nokes-Malach, 2023).
  - **EOL:** 학습 전 학습 난이도 판단 ➔ **문제를 보기만 하고 풀 수 있을지 미리 추정하는 판단**(Burton et al., 2017).
  - **JOL:** 항목에 대해 학습 중에 얼만큼 확신을 하면서 학습중인지 판단 ➔ 문제 풀이 중 직관을 지속적으로 **모니터링**하여 해결 중인 **문제에 대한 확신도**를 판단(Metcalfe & Wiebe, 1987).
  - **FOK:** 회상 실패 직후 추후에 단서를 주면 기억할 수 있을지 판단 ➔ 지금은 못 풀었지만, **단서나 추가 정보**가 주어지면 나중에 풀 수 있을지에 대한 예측(Metcalfe, 1986).

<표 2> 본 연구에서 정의한 메타인지 각 단계별 정의

단계	정의
EOL (Ease-of-Learning Judgements)	**질의 처리 용의성 판단**: 받은 질의가 얼마나 답변하기 쉬울지 사전 **평가**
JOL (Judgemets-of-Learning)	**현재 답변 완성도 판단:** 답변을 얼마나 확신하면서 생성했는지를 **평가**
FOK (Feeling-of-Knowing)	**단서 활용 적합도 판단:** 기존에 주어진 단서(문서 검색)으로 충분히 잘 답변하였는지 **평가**

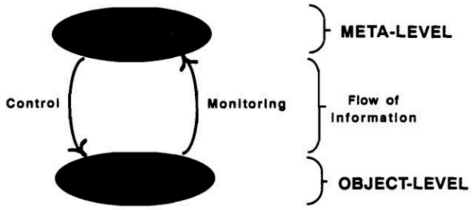
- ***본 연구에서는 위 메타인지 단계에서 정의한 평가를 수행하기 위해 각 단계별 지표를 선정하여 단계별 점수를 정량화 함.***
- ***또한, 단계별 점수가 낮은 단계에 대해 개입해 산출된 지표 점수를 고려해서 쿼리를 재작성하여 답변 품질을 향상시키고자 함.***

#3. 연구 방법 (Research Methods)

: 3-1 Nelson & Narens(1990)와 본 연구의 대응

3.1 Nelson & Narens(1990)와 본 연구의 대응

- Nelson & Narens (1990)은 인간 학습을 **객체-수준과 메타-수준으로 분리하고, 두 수준이 상향 보고·하향 제어 루프를 이룰 때 학습 효율이 최적화**된다고 설명함.
- **본 연구는 이 이중 수준 구조를 기능적 유사성 차원에서 대화형 AI에 적용함.**
  - **객체 수준**
    대화형 AI는 질의에 대한 **응답을 생성**하며, 동시에 텍스트 생성 과정에서 EOL, JOL, FOK와 대응되는 **메타인지 지표**를 산출함
  - **메타 수준**
    추출된 메타인지 신호는 경량 분류기에 입력되며, 이 모듈은 meta score를 계산하여 **질의 재작성 여부**를 결정함.
  - **상향 보고**
    객체 수준에서 산출된 메타인지 지표가 메타 수준으로 전달됨
  - **하향 제어**
    메타 수준에서 산출된 최종 의사결정을 통해 객체 수준의 행동이 실제로 조절됨



<그림 6> 객체-메타 수준 프레임워크(Nelson & Narens, 1990)

#3. 연구 방법 (Research Methods)

: 3-2 MTRAG 데이터셋

3.2 MTRAG 데이터셋

- MTRAG는 멀티턴 RAG 대화 평가를 위한 최초의 종단간 인간 생성 벤치마크 데이터셋으로, 실제 세계의 대화를 반영하여 RAG 파이프라인 전체를 평가할 수 있는 다차원 접근법을 제공함.
- **평균 대화 당 7.7턴의 데이터를 구성**하고 있으며 ClapNQ(일반 지식), FiQA(금융), Govt(정부 지식), Cloud(IT)로 구성된 4개의 도메인을 확보하고 있음.
- 평가 방법으로는 **RAGAS 지표**(Faithfulness, Context Precision/Recall, Answer Relevancy)를 적용하여 자동화된 성능 평가를 수행해서 **검색·생성**의 성능을 측정함.



<그림 7> ragas score 평가 프레임워크(Es et al., 2024)

#3. 연구 방법 (Research Methods)

: 3-3 평가 지표

3.3 평가 지표

<표 3> 평가 지표

항목	지표	설명
generation	RAGAS – Answer Faithfulness	응답이 근거 문맥과 사실적으로 일치하는 정도
	RAGAS – Answer Relevance	응답이 질문과 얼마나 관련있는지 보는 지표
retrieval	RAGAS – Context Precision	참조된 문맥 중 실제로 필요한 문맥의 비율
	RAGAS – Context Recall	필요한 문맥 중 실제로 활용된 문맥의 비율

- RAGAS는 RAG와 LLM 애플리케이션의 품질을 자동으로 평가하는 오픈소스 라이브러리이자 프레임워크로써, 정답 라벨 없이도 평가가 가능하도록 설계된 방식의 지표들을 제공함.
- **검색과 생성** 모듈을 분리해 각 단계의 성능을 따로 측정하는 지표 세트를 제공함.
- 기존 BLEU, ROUGE, EM처럼 표면적 유사도만 보는 지표와 달리, **문맥 근거와 질문 관련성, 회수 품질**을 함께 다루므로 더 다차원적인 답변 품질 측정이 가능함.

#3. 연구 방법 (Research Methods)

: 3-4 연구 설계

3.4 연구 설계

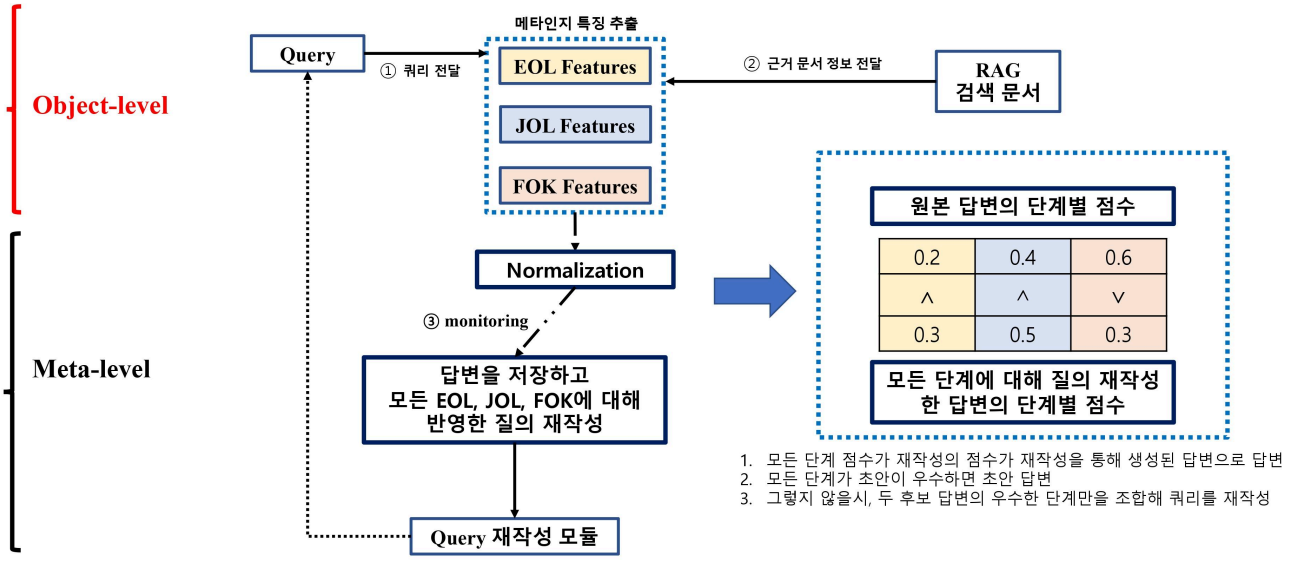
<표 4> 각 단계별 특징 지표

단계	변수명	기호	설명	채택 근거
EOL	희귀 토큰 비율	R_{rare}	질문에 등장한 희귀/생소 토큰 비율	질의의 어휘 분포가 산만하면 모델의 이해 성능이 떨어지며, 질의의 '명료도'같은 사전 지표가 실제 평균정확도와 유의하게 상관함
	구문 복잡도	C_{com}	길이·절/접속사 기반 구조 복잡도	의존거리·작용기억 비용이 큰 문장은 처리 난도가 올라가고 오류가 늘어날 수 있음
JOL	로짓 마진 평균	$\bar{M}$	각 토큰에서 1등과 2등의 로짓 차의 평균	마진이 클수록 모델이 정답에 더 확신하고 작을수록 모델이 여러 선택지 사이에서 헷갈리고 있다는 신호
	엔트로피 확신도	C_H	분포 엔트로피 기반 확신도	토큰/시퀀스 불확실성(엔트로피)은 오류·환각과 정적으로 연결되어 있음
	윈도우형 최근 평균 NLL	L_{win}	가까운 토큰들에 대한 Negative Log-Likelihood들의 평균	텍스트의 난이도가 변하는 구간을 감지할 수 있고 따라서 스트리밍 생성 중 모델의 성능을 추적할 수 있음
FOK	쿼리-문서 정렬도	$S_{qd}@k$	질의와 상위 k 문서 사이의 의미 정렬 정도	근거 검색 오정렬과 질의 모호성을 가장 먼저 드러내며 재검색이나 질의 재작성의 필요성을 드러낼 수 있음
	문서-답변 근거화 점수	S_{DA}	최종 답변이 인용된 근거 스팬에 의해 지지되는 정도로 산출 방식 답변과 인용 스팬의 의미 유사도로 검증	채택 근거 생성이 근거에 정렬되는지 바로 반영하므로 환각 위험과 근거 미흡을 식별할 수 있음
	지표명 근거 일관성 지수	$C_{DD}@k$	상위 k 문서들 사이의 내용 합치 정도를 고려한 일관성 수준	검색 문서가 서로 충돌할 때 발생하는 혼탁과 오류 전파를 측정하여 필터링과 재분배에 즉시 활용 가능함

#3. 연구 방법 (Research Methods)

: 3-5 Research Framework

3.5 연구 프레임워크

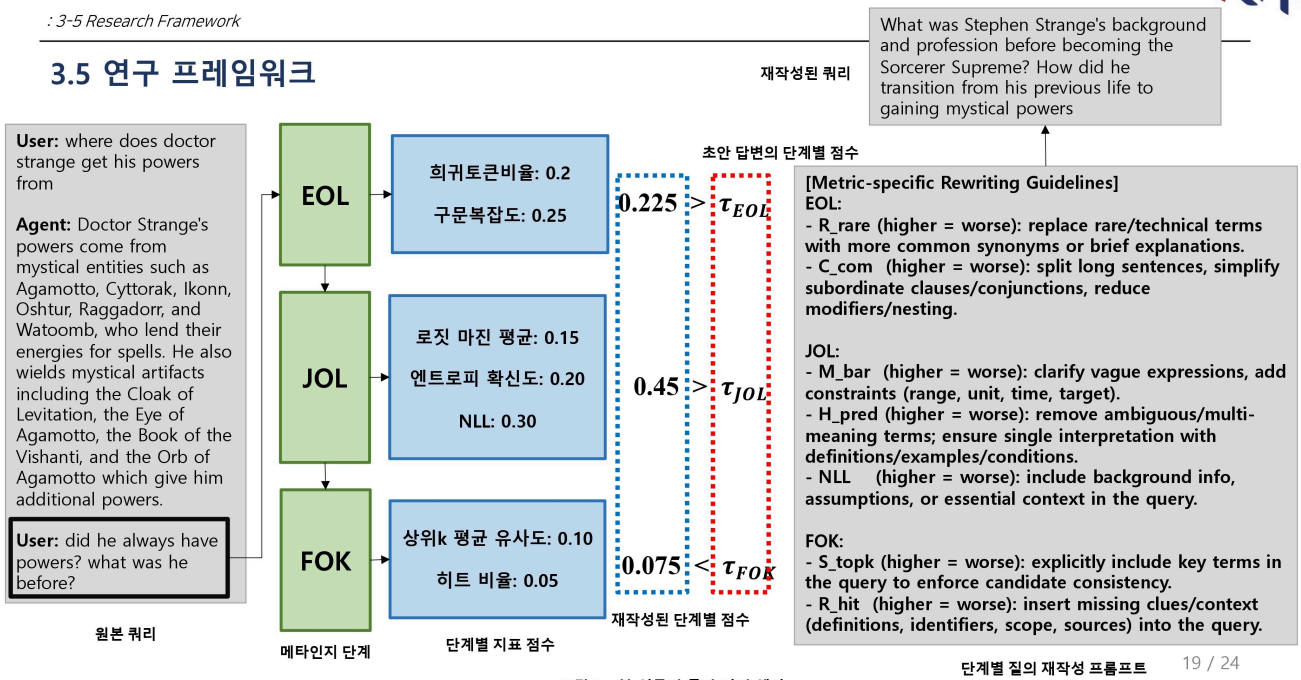


<그림 8> 본 연구의 프레임워크

#3. 연구 방법 (Research Methods)

: 3-5 Research Framework

3.5 연구 프레임워크



<그림 9> 본 연구의 동작 과정 예시

#3. 연구 방법 (Research Methods)

: 3-6 질의 재작성 프롬프트

3.6 질의 재작성 프롬프트(개발중)

EXAMPLE

```
You are a "Metacognition-based Query Rewriter".
The input is always JSON, containing the original query, stage scores (EOL/JOL/FOK), and stage thresholds only.
No metric-level thresholds are given.
Your task is to apply the decision rules and metric-specific rewriting guidelines below.
For every stage whose score exceeds its threshold, you must perform at least one rewrite.
Return the result strictly as JSON.

[Input JSON Schema]
{
  "query": "<original query>",
  "stages": {
    "EOL": {
      "value": <float>,          // 0~1 (0=almost no issue, 1=severe issue)
      "threshold": <float>,      // stage threshold θ_EOL
      "metrics": {               // metric values only (no thresholds)
        "R_rare": <float>,       // rare-token ratio
        "C_com":  <float>        // syntactic complexity
      }
    },
    "JOL": {
      "value": <float>,
      "threshold": <float>,      // θ_JOL
      "metrics": {
        "M_bar":  <float>,       // logit margin mean
        "H_pred": <float>,       // predictive entropy
        "NLL":    <float>        // negative log-likelihood
      }
    },
    "FOK": {
      "value": <float>,
      "threshold": <float>,      // θ_FOK
      "metrics": {
        "S_topk": <float>,       // top-k mean similarity
        "R_hit":  <float>        // hit ratio
      }
    }
  },

  // (Optional) metric selection parameters; defaults apply if absent
  "selection": {
    "mode": "by_stage|top_k|top_p",  // default: "by_stage"
    "k": 1,                          // used if mode=top_k
    "p": 0.5                         // used if mode=top_p (proportion 0~1)
  }
}
[Decision Rules]
1) Stage activation:
   - If stage value ≥ stage threshold, mark the stage as activated=true.
   - Otherwise, mark as deactivated.

2) Metric selection (no metric thresholds):
   - For an activated stage, choose metrics according to the selection mode:
   - Default (mode="by_stage"): select all metrics with value ≥ the stage's value.
     Guarantee rule: if no metrics qualify, select the single highest-value metric (top-1).
   - mode="top_k": select the top-k metrics by value (limited to available metrics).
   - mode="top_p": sort metrics descending, select the top p proportion (at least one metric).
```


<그림 10> 본 연구의 질의 재작성 프롬프트

#4. 예상되는 분석 결과 (Expected Analysis)

: 4-1 예상되는 분석 결과

4.1 예상되는 분석 결과

<표 5> 예상되는 성능표

모델	RAGAS-Answer Faithfulness	RAGAS – Answer Relevance	RAGAS – Context Precision	RAGAS – Context Recall
*Llama 3.1 8B	0.847918	0.457570	0.559077	0.478674
**Llama 3.1 8B				
***Llama 3.1 8B				
*Qwen 2.5 7B				
**Qwen 2.5 7B				
***Qwen 2.5 7B				

- *는 질의 재작성을 하지 않은 원본 모델
- **는 모든 질의에 대해 재작성을 한 모델
- ***는 본 연구가 제안한 메타인지 기반 질의 재작성모델

#5. 예상되는 결론 (Expected Conclusion)

: 5-1,2 학술적/실무적 기대효과

5.1 학술적 기대효과

- Nelson & Narens(1990)의 메타인지를 대화형 AI 문제해결에 기능적으로 매핑한 프레임워크를 제시함.
- 각 단계별 지표에 대응하는 구체 지표군을 정의하여 전 주기적 모니터링에 활용될 수 있는 방안을 제시함.
- 멀티턴 맥락에서 답변 품질 향상을 다루기 위해 진단, 개입, 검증 프로토콜과 평가 설계를 제시해 후속 연구의 공통 기반을 제공함.

5.2 실무적 기대효과

- 무차별 재작성 대신 취약 성분만 선택적으로 보강하여 대화형 AI의 답변 품질을 안정화함.
- 단계별 지표를 투명하게 수집하여 오류 원인을 파악하고 운영 품질 관리에 기여할 수 있음.
- 다양한 에이전트·도메인에 쉽게 이식이 가능하여 실서비스에 적용할 수 있음.

2025 한국지식경영학회 추계학술대회

An Empirical Study on Input Modality Combinations for Improving Visual Understanding in Multimodal Large Language Models

멀티모달 대규모 언어 모델의 시각 이해 개선을 위한 입력 모달리티 조합에 대한 실증 연구

김 도 원[1] · 조 윤 호[2] · 문 현 실[3]
Do Won Kim | Yoon Ho Cho | Hyun Sil Moon

1) 국민대학교 경영대학원 | riseup7@naver.com
2) 국민대학교 경영대학 | www4u@kookmin.ac.kr
3) 국민대학교 경영대학원 | hsmoon@kookmin.ac.kr

목차

1. 서론
2. 관련 연구
3. 연구 방법
4. 실험 결과
5. 결론

서론

❍ Large Language Model (LLM)

- 2022년말 ChatGPT의 등장은 LLM의 대중화를 이끌었고 다양한 분야에서 활용 중 (Ouyang et al., 2022)
  1) 일반 사용자들도 뛰어난 접근성 바탕으로 손쉽게 인공지능 기술을 활용 가능
  2) 탁월한 텍스트 이해 및 생성 능력 바탕으로 의료·산업 ·금융 등 다양한 분야에서 활용 중 (Urlana et al., 2024; Wang et al., 2024; Zhao et al., 2024)
- LLM의 한계 : 실제 세계는 시각·청각·공간 등 다양한 정보가 있어 자연어 처리만으로는 활용 범위에 한계 존재

❍ Multimodal Large Language Model (MLLM)

- 텍스트, 이미지, 소리 및 영상 등 다양한 모달리티를 통합적으로 처리할 수 있는 MLLM으로 발전
- 언어 중심의 정보 처리에서 시각적 맥락을 함께 이해하고 활용하는 방향으로 진화 (Yin et al., 2024)
- 특히, 시각 정보는 높은 함축성과 직관성 바탕으로 효율적인 의사소통 가능 (Henning & Ewerth, 2017)
- MLLM의 한계 : 이미지 인식 능력은 우수하나, 시각 정보를 이해하고 추론하는 능력은 미흡 (Małkiński et al., 2024)

서론

❍ 시각적 추론의 정의

- 시각적 추론이란 인식된 정보를 기반으로 관계·맥락·상식을 고려해 논리적 결론을 도출하는 고차원적 이해과정 (Zellers et al., 2019)

  *참고) 시각 인식 : 이미지나 영상에서 객체·속성·장면을 식별하고 분류하는 과정 (Luo et al., 2023)

❍ MLLM의 시각 추론 능력

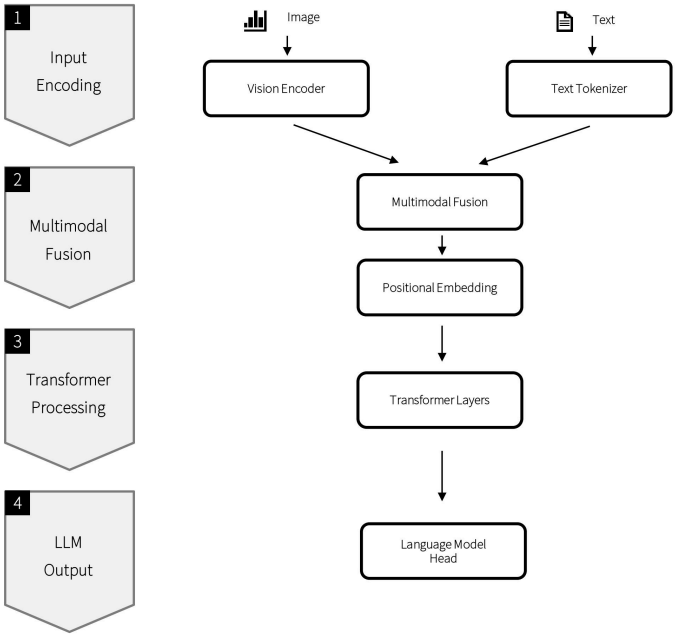
- 텍스트가 포함된 시각 정보(포스터·도표·문서)의 의미 해석 실패 (Hu et al., 2024)
- 최신 모델 포함 15종의 MLLM의 시각 추론 과제로 평가했을 때 전 모델 정확도 60% 미만을 기록 (Yao et al., 2025)
- 언어적 사고에 편향되어 있어 시각 정보를 구조적으로 해석하는 데 한계가 있음 (Hao et al., 2025)
- 인식(Recognition)에서 지식 기반 인지(Cognition) 단계로의 확장 필요성 제기 (Zellers et al., 2019)

Research Purpose

MLLM의 시각적 추론 능력 향상을 위해 입력 형태와 조합별 성능 차이를 실증적으로 분석하고,
그 결과를 바탕으로 최적 멀티모달 입력 전략을 제안

관련 연구

❍ MLLM 의 시각 정보 처리 과정



1 Input Encoding

- Vision Encoder : 이미지 입력을 받아 고차원 특징 벡터로 변환
- Text Tokenizer : 텍스트를 토큰 시퀀스로 분해하고 임베딩 벡터로 변환

2 Multimodal Fusion

- Multimodal Fusion : 이미지와 텍스트 특징을 하나의 통합된 멀티모달 표현으로 결합
- Positional Embedding : 토큰의 순서 정보를 벡터에 추가

3 Transformer Processing

- Transformer Layers : Self-Attention 메커니즘을 통해 결합된 멀티모달 임베딩 벡터들 간의 관계와 연결된 의미를 분석하고 이미지-텍스트 상호작용 처리

4 LLM Output

- Language Model Head : 토큰 확률 분포를 계산하여 다음 출력 생성

참고:Team G. et al. (2025); Bai et al. (2025)

관련 연구

❍ MLLM의 시각 추론 연구의 발전 과정

Phase 1. 객체 인식

- 시각 정보 처리의 기초 단계
- 이미지 내 객체의 존재와 위치 파악

Phase 2. 관계 인식

- 시각 정보를 심층적으로 이해하는 단계
- 객체 간 상호작용 및 관계 추론

Phase 3. Attention 메커니즘의 도입

- 유의미한 정보에 선택적 집중
- Vision Transformer(ViT)의 등장 (Bahdanau, 2014)

❍ MLLM의 시각 추론 한계 분석 연구

- MLLM의 언어 중심적 편향 사고 (Hao et al., 2025)
  1) 시각 추론 벤치마크(EMMA)에서 최고 성능 모델(o1)의 정확도는 45.8%로 사람 추론 능력과 24.2% 성능 격차 발생
  2) 오답의 52.8%가 시각적 관계 이해 실패·공간적 변환 오류에서 발생
  3) 추론 시 텍스트 정보에 대한 의존도가 높으며, 시각 정보를 기반으로 추론하는 능력 부족
- MLLM 시각 추론 성능 저하의 구조적 요인 (Tong et al., 2024)
  1) 세부적인 시각 정보 차이 판별 실험에서 인간 대비 GPT-4V, Gemini 모델 정확도 58~60% 성능 격차 발생
  2) 시각 정보를 처리하는 비전 인코더가 이미지의 전체적인 의미 파악에 집중되어 세부적인 인식 및 변별력 부족

연구방법

❍ 실험 설계

Ⅰ. 데이터 준비 및 전처리

1 실험 대상 기업 선정

- 2024년 기준 KOSPI 시가총액 상위 50개 기업 선정

2 데이터 수집 및 전처리

- 2021 ~ 2024년 OHLCV(Open, High, Low, Close, Volume) 주가 데이터 수집
- 20일씩 그룹화하여 입력 시퀀스 구성 후 1일씩 이동하며 샘플 생성 (Sliding Window 방식)
- 캔들 차트 이미지, 이미지 해석 데이터 생성

3 입력 모달리티 구성

- 5가지 입력 조합 구성하여 실험 수행

Ⅱ. 모델 미세 조정

1 실험 대상 모델 선정

- Qwen-2.5-VL-32B
- Gemma-3-27B

2 QLoRA 미세 조정

- 기존 언어 모델 4-bit 양자화 후 가중치 고정
- Train data set을 이용하여 어댑터 LoRA 어댑터 미세 조정
- Loss Function 계산 후 어댑터 가중치 지속 업데이트

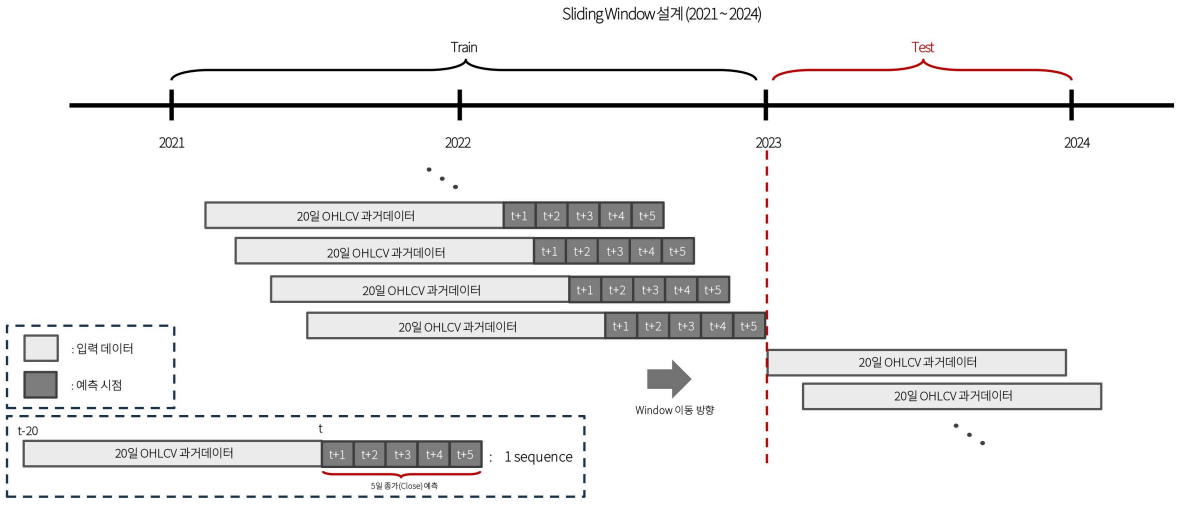
3 모델 평가

- RMSE 지표를 이용하여 모델 예측값 평가
- 5일 예측 종가와 실제 종가 간 오차 측정

연구방법

❍ 데이터 준비 및 전처리

1. 시계열 데이터 예측에서 자주 이용되는 Sliding Window 방식으로 데이터를 구성
2. 연속된 과거 20일치 주가 데이터(t-20 ~ t)와 직후 5영업일 종가 데이터(t+1 ~ t+5)를 하나의 시퀀스로 구성하여 2021 ~ 2024 기간 데이터 생성
3. 2021 ~ 2023년 3년치를 Train, 2023 ~ 2024년 1년치를 Test 데이터 셋으로 분할



연구방법

❍ 데이터 준비 및 전처리

입력 모달리티 조합 구성

1 정형 데이터

- 각 기업의 OHLCV(Open, High, Low, Close, Volume) 주가 데이터 수집
- 20일치 OHLCV 주가 데이터 시퀀스를 구조화된 텍스트로 구성하여 JSON 형식으로 나열

2 이미지

- 각 정형 데이터 시퀀스를 캔들 차트 이미지로 생성
- MLLM의 순수 시각적 추론 능력을 평가하기 위한 목적

3 이미지 + 정형 데이터

- 캔들 차트 이미지와 원천 정형 데이터를 같이 제공
- 차트와 원천 수치 데이터를 구조화된 형태로 함께 제공하여 발생하는 효과 평가

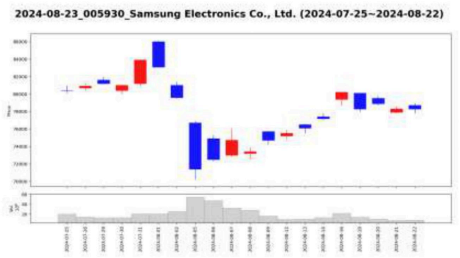
4 이미지 + 이미지 해석 데이터

- 이미지를 사전 학습된 MLLM에 제공하여 차트 해석 데이터를 생성
- 이미지 해석 데이터를 같이 제공함으로써 언어적 맥락이 시각 추론 향상에 미치는 영향 평가

5 이미지 + 정형 데이터 + 이미지 해석 데이터

- 모든 모달리티를 통합 제공하여 추론 능력 극대화 가능성 검증

입력 데이터 예시



< 이미지 >

"input": {"target_date": "2024-01-02",
"ticker": "005930",
"name": "Samsung Electronics Co., Ltd.",
"historical_data(OHLCV)": "2023-11-30 72700 72800 72200 72800 15783714\n2023-12-01 72400 72500 71700 72000 9871284\n2023-12-04 72800 72900 72400 72600 10229267\n2023-12-05 72300 72400 71200 71200 12129682\n2023-12-06 71800 72100 71600 71700 8123087\n2023-12-07 71800 71900 71100 71500 8862017\n 2023-12-08 72100 72800 71900 72600 10859463\n2023-12-11 72800 73000 72200 73000 9861960\n · · · "}

< 정형 데이터 >

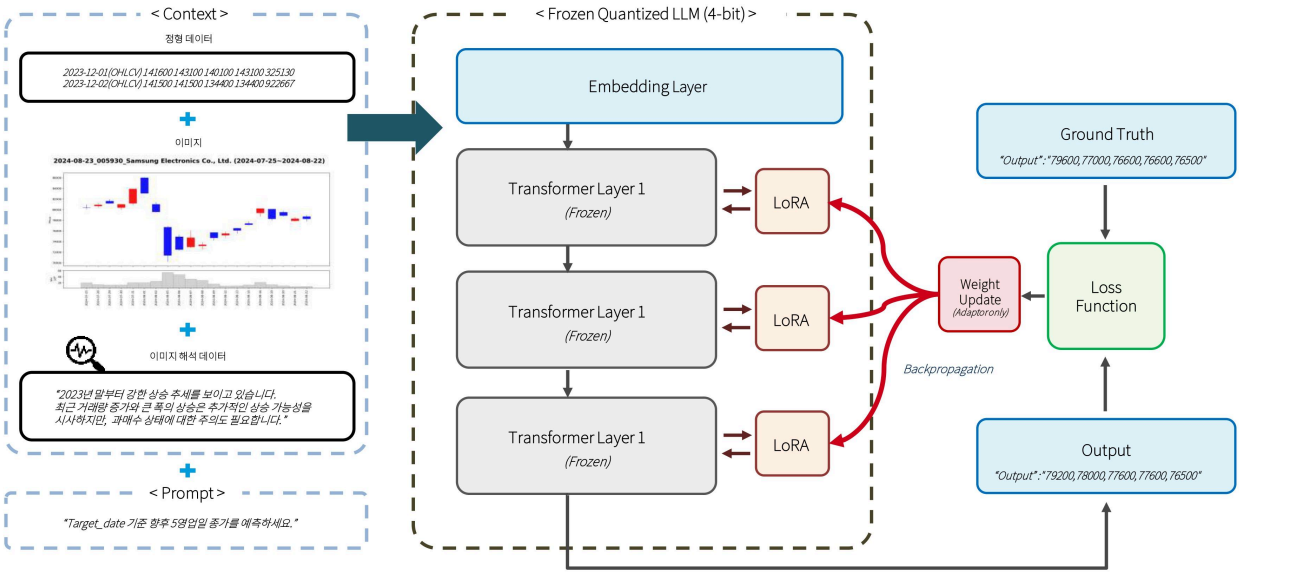
"input": {"target_date": "2024-01-02",
"ticker": "005930",
"name": "Samsung Electronics Co., Ltd.",
"interpretation": "'Pattern': '11월 말 하락세로 시작하여 12월 중순 강한 상승세를 보였으나, 12월 말 다시 하락하며 변동성이 증가했습니다.',
'Supplement': '12월 14일과 28일에 거래량이 급증하며 각각 상승 전환점과 하락 시점에서 중요한 거래가 발생했습니다.',
'Summary': '12월 중순까지의 상승 추세가 최근 하락세로 전환되었으며, 단기 조정 또는 하락 추세 시작 가능성이 있지만,
동시에 추가 상승 가능성도 존재하여 주의가 필요합니다."}

< 이미지 해석 데이터 >

연구방법

❍ 모델 미세 조정

【 QLoRA 미세 조정 구조 】



참고 : Dettmers et al. (2025)

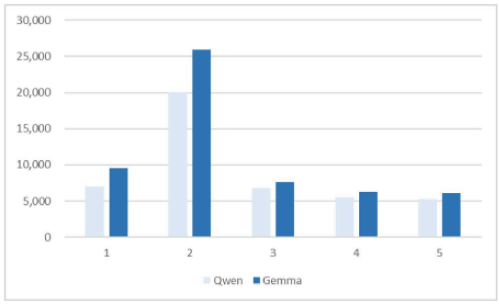
실험 결과

【 RQ 1. 모델 구조에 따라 성능 차이가 발생하는가? 】

Set	Input	Mean(Qwen)	Mean(Gemma)	Mean Diff	t-value	Effect size (Cohen's d)
1	정형 데이터	7,007.72	9,504.28	2,496.56	23.19***	0.21
2	이미지	20,035.14	25,891.47	5,856.33	71.78***	0.65
3	이미지 + 정형 데이터	6,799.56	7,622.30	822.74	75.13***	0.68
4	이미지 + 이미지 해석 데이터	5,542.98	6,293.39	750.40	81.77***	0.74
5	이미지 + 정형 데이터 + 이미지 해석 데이터	5,291.66	6,102.71	811.05	59.50***	0.54

*** : p-value < 0.001, ** : p-value < 0.01, * : p-value < 0.05

모델별 성능 시각화



결과 해석

1. 전반적으로 Qwen의 성능이 Gemma보다 평균 약 2,147 RMSE만큼 우수함

2. Set 2에서 약 22.6% 성능차이를 보이며 격차가 가장 크게 나타남

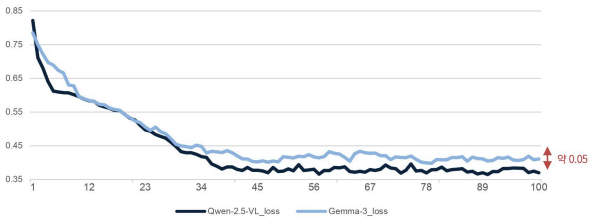
ㄴ해석 : 순수 시각 정보 처리에서 Qwen이 Gemma보다 우수함

3. Set 4에서 효과 크기가 가장 크게 나오며, 해당 조합에서 차이가 가장 두드러지게 나타남

ㄴ해석 : 언어적 맥락 정보가 제공될 때 모델 구조적 특성에 따른 성능 차이가 가장 큼

실험 결과

❍ 모델별 Epoch에 따른 Validation Loss 추세



학습 수렴 속도

- 초기 구간에서 Qwen모델은 Gemma보다 Loss 감소율이 높음

→ Qwen이 학습 속도가 더 빠르며, 이는 파라미터 규모 차이로 인한 학습 수렴 속도 우위

최종 Loss 수렴 값의 차이

- 최종 Loss 수렴값은 Qwen 모델이 Gemma보다 약 5% 낮은 값을 기록

→ 이는 Qwen이 시각 정보 이해 측면에서 구조적으로 우수함을 시사함

❍ 모델 구조 차이 분석

	비전 인코더	모달 융합 전략	위치 임베딩
Qwen	• ViT - Dynamic Resolution	• Vision-Language Adapter - 시각 정보를 언어 공간에 정렬하여 통합 - Semantic Fusion (의미적 융합)	• M-RoPE (Multimodal Rotary Position Embedding) - 시간/높이/너비 3축 기반 통합 위치 인코딩 - 텍스트와 이미지의 위치 정보를 동일 좌표계에서 처리
Gemma	• SigLIP - Fixed Resolution	• No Adapter - Direct Concatenation	• Standard RoPE (Rotary Position Embedding) - 텍스트 토큰의 순서 정보만 인코딩 - 이미지 패치 간 공간 관계 정보 미반영

참고 : Bai et al. (2025); Heo et al. (2024); Liu et al. (2024); Team G. et al. (2025); Wang et al. (2024)

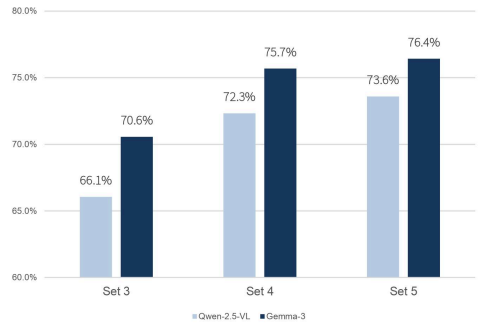
실험 결과

【 RQ 2. 입력 모달리티 조합에 따라 MLLM의 시각적 추론 성능은 어떻게 달라지는가? 】

Set	Input	Mean(Qwen)	Mean(Gemma)	Mean Diff	t-value	Effect size (Cohen's d)
1	정형 데이터	7,007.72	9,504.28	2,496.56	23.19***	0.21
2	이미지	20,035.14	25,891.47	5,856.33	71.78***	0.65
3	이미지 + 정형 데이터	6,799.56	7,622.30	822.74	75.13***	0.68
4	이미지 + 이미지 해석 데이터	5,542.98	6,293.39	750.40	81.77***	0.74
5	이미지 + 정형 데이터 + 이미지 해석 데이터	5,291.66	6,102.71	811.05	59.50***	0.54

*** : p-value < 0.001, ** : p-value < 0.01, * : p-value < 0.05

데이터 결합에 따른 성능 향상률



결과 해석

1. 모든 모달리티 통합의 효과 (Set 5)

: 모든 모달리티가 통합될 때 시각적 추론 능력이 극대화됨을 확인

2. 언어적 맥락의 효과성 검증 (Set 4)

: 이미지 해석 데이터가 정형 데이터보다 성능 향상에 더 크게 기여

└해석 : 시각 정보에 대한 언어적 맥락 설명이 정형 데이터보다 시각적 추론 능력 향상에 유리

3. 시각 정보 처리 능력의 한계 (Set 1 vs Set 2)

: 정형 데이터 단독 입력이 이미지 단독 입력보다 RMSE 성능이 우수함

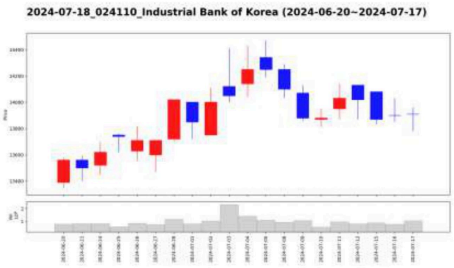
4. 모달리티 통합 개선률의 한계 (Set 4 vs Set 5)

: 모든 정보를 결합한 Set 5의 개선 폭은 Set 4 대비 약 4.5%로 제한적임

└해석 : 정보의 중복성으로 인해 추가적인 성능 향상이 크지 않음

실험 결과

【 RQ 3. 이미지의 시각 단서 중 MLLM의 추론 성능을 향상시키는 것은 무엇인가? 】



< Good case >



< Bad case >

우수 사례

- 양봉과 음봉이 혼재된 복합적인 패턴이 나타남
- 상승과 조정이 반복되는 다양한 변동성
- 풍부한 맥락 정보 제공 가능

저조 사례

- 양봉과 음봉이 연속적으로 나오며 단조로운 패턴이 나타남
- 단순한 패턴에 모델이 과적합 되는 현상 발생
- 제한된 정보로 인해 일반화 능력 저하

✓ **복합적이고 다양한 시각적 패턴이 MLLM의 시각적 추론 능력을 향상시키며, 단조로운 패턴은 오히려 모델의 과적합을 유발**

결론

결과 요약

1. 멀티모달 통합 입력 전략의 효과
- 모든 모달리티 통합 시 최고 성능 달성

2. 언어적 맥락 정보의 중요성
- 이미지 해석 데이터 결합 시 72.3% 성능 개선
- 정형 데이터 결합보다 높은 기여

3. 모델의 구조적 차이의 영향
- Qwen-2.5-VL이 Gemma-3 대비 평균 12% 성능 우수

4. 시각적 패턴 복잡도의 영향
- 복합적 패턴이 단조로운 패턴보다 시각적 추론 성능 우수

연구 기여

1. 입력 데이터 전략의 중요성 실증
- 시각적 추론 개선을 위한 효과적인 입력 전략 제안

2. 언어적 맥락의 효과 규명
- 정형 데이터보다 시각 정보에 대한 언어적 맥락 정보의 중요성 입증
- MLLM의 시각 추론 메커니즘에 대한 기반 지식 확장

3. 모델 내부 구조의 중요성 제고
- 비전 인코더, 멀티모달 융합 전략, 위치 임베딩 방식의 중요성 입증

4. 시각적 추론 능력의 한계 및 개선 방향 제시
- 다양한 시각적 복잡도에 강건한 모델 개발 필요

한계 및 향후 연구 방향

한계
- 실험이 금융 도메인의 캔들 차트에 국한되어 있어 본 연구 결과의 일반화 가능성에 제약 존재
- 오픈소스 대표적 모델인 Qwen과 Gemma모델만 비교하여 다양한 MLLM에 대한 포괄적 분석 부족

향후 연구 방향
- 다양한 도메인의 시각 정보를 대상으로 멀티모달 입력 전략의 효과 검증
  - 예시 : 의료 이미지, 산업 도메인, 교육 자료, 지리 정보 등
- 다양한 MLLM 구조를 비교하여 최적 입력 전략 도출 및 표준화
  - 예시 : 비전 인코더(CLIP, DINOv2, Flamingo, BLIP-2), 위치 임베딩 방식(ALiBi) 등

Q & A

참고문헌

1. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., ... & Lin, J. (2025). Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*.
2. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
3. Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems*, 36, 10088-10115.
4. Hao, Y., Gu, J., Wang, H. W., Li, L., Yang, Z., Wang, L., & Cheng, Y. (2025). Can MLLMs reason in multimodality? EMMA. *arXiv preprint arXiv:2501.05444*.
5. Henning, C. A., & Ewerth, R. (2017). Estimating the information gap between textual and visual representations. *In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, 14-22.
6. Heo, B., Park, S., Han, D., & Yun, S. (2024). Rotary position embedding for vision transformer. *In European Conference on Computer Vision*, 289-305
7. Hu, W., Xu, Y., Li, Y., Li, W., Chen, Z., & Tu, Z. (2024). Bliva: A simple multimodal llm for better handling of text-rich visual questions. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 38(3), 2256-2264.
8. Liu, W., Zhu, F., Ma, S., & Liu, C. L. (2024). MSPE: multi-scale patch embedding prompts vision transformers to any resolution. *Advances in Neural Information Processing Systems*, 37, 29191-29212.
9. Luo, S. et al. (2023). *Towards a Unified Neural Architecture for Visual Recognition and Reasoning*. arXiv:2311.06386.
10. Małkiński, M., Pawlonka, S., & Mańdziuk, J. (2024). Reasoning limitations of multimodal large language models**.** *arXiv preprint arXiv:2411.01173*.
11. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35, 27730-27744.
12. Team G., Kamath, A., Ferret, J., Pathak, S., Vieillard, N., Merhej, R., ... & Iqbal, S. (2025). Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*.
13. Tong, S., Liu, Z., Zhai, Y., Ma, Y., LeCun, Y., & Xie, S. (2024). Eyes wide shut? exploring the visual shortcomings of multimodal llms. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9568-9578.
14. Urlana, A., Kumar, C. V., Singh, A. K., Garlapati, B. M., Chalamala, S. R., & Mishra, R. (2024). LLMs with industrial lens. *arXiv preprint arXiv:2402.14558*.
15. Wang, D., & Zhang, S. (2024). Large language models in medical and healthcare fields: applications, advances, and challenges. *Artificial intelligence review*, 57(11), 299.

참고문헌

16. Wang, P., Bai, S., Tan, S., Wang, S., Fan, Z., Bai, J., ... & Lin, J. (2024). Qwen2-VL. *arXiv preprint arXiv:2409.12191*.
17. Yao, R., Zhang, B., Huang, J., Long, X., Zhang, Y., Zou, T., ... & Van Gool, L. (2025). LENS. *arXiv preprint arXiv:2505.15616*.
18. Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024). A survey on multimodal large language models. *National Science Review*, 11(12).
19. Zellers, R., Bisk, Y., Farhadi, A., & Choi, Y. (2019). From recognition to cognition: Visual commonsense reasoning. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6720–6731.
20. Zhao, H., Liu, Z., Wu, Z., Li, Y., Yang, T., Shu, P., ... & Liu, T. (2024). Revolutionizing finance with LLMs. *arXiv preprint arXiv:2401.11641*.

	Memo	

디지털 기업 창업자와 투자자 간 ESG 경영요소 도입 우선순위 인식 비교에 관한 연구 : 혼합 방법론 사용

전남대학교 경영학부
박세민, 고준

목차 table of contents

1 연구 배경과 목적
2 이론적 배경
3 질적 연구
4 양적 연구
5 결론



Part 1 연구 배경과 목적

□ 정부와 민간의 노력으로 최근 제 2의 벤처 붐이 본격 도래

- 중기부의 창업지원 예산은 98년 최초 82억 원을 시작으로 24년 약 3조 7천억원으로 첫 예산 편성 대비 29,600배 증가 (중기부, 2024)

□ 매년 10만개 이상 신규창업이 발생하나 창업 생존율은 5년 차에 30%에 그쳐

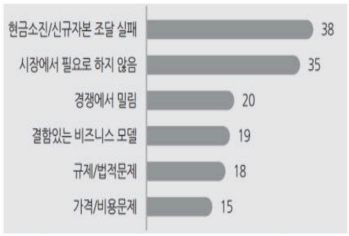
□ 창업 실패의 가장 큰 원인은 신규 자금조달 실패(38%) (CB Insights, 2021)

- 자금 조달의 주체는 국가(정부 및 지자체, 창업지원기관의 창업지원금) 및 민간투자사(액셀러레이터 및 벤처캐피털의 투자 자금)
- 창업 1~3년 차는 정부지원금으로 생존이 가능하나 3년 차 이후부터는 민간 투자사의 투자가 필수

[창업기업 생존율]



[창업 실패 원인]



※ 자료 : OECD, 중소벤처기업부, CB Insights

Part 1 연구 배경과 목적

□ 민간 투자사의 투자 결정요소는 전통적으로 재무적 성과를 중시하였으나 현대에는 장기적 기업가치에 영향을 미치는 비재무적 요소(ESG 경영)도 함께 고려

- ESG 경영전략이 사회친화적 기업 이미지를 구축하여 소비자 및 투자자로부터 긍정적 행위를 이끌어내 기업가치 향상에 긍정적 영향을 주기 때문 (허종오 외, 2022)

□ 대기업에 비해 초기 창업기업은 기술개발, 판로개척, 매출 확보 등 단기적 성과에 집중하게 되어있어 장기적 전략인 ESG 경영을 실천하기 어려운 실정

- 투자 결정 시 ESG를 중요하게 생각하는 투자자와 투자유치가 반드시 필요하지만 ESG 경영도입이 어려운 창업자 간의 괴리

⇒ 본 연구는 창업자와 투자자가 인식하는 ESG 경영요소 도입 우선순위 인식의 차이를 비교·분석하고자 함

창업자와 투자자의 인식 격차를 규명하고 이에 따른 실무적 시사점을 도출
창업자 : 투자자 관점의 ESG 요소 우선순위를 인식하고 지속 성장을 위한 투자유치 및 ESG 경영 전략 수립에 도움
투자자 : 창업자의 현실적 여건과 ESG 도입 인식을 반영한 투자 판단 기준 제공

Part 2 이론적 배경 (액셀러레이터 및 벤처캐피털)

액셀러레이터

□ 초기 스타트업을 발굴하여 투자 및 창업 공간, 멘토링 등 종합적인 육성 서비스를 제공하는 창업촉진 전문회사 및 기관을 뜻 함(이정우, 2016)

- 2005년 미국에서 설립되어 에어비앤비 등 다수의 유니콘 기업을 키워낸 최초의 액셀러레이터 Y 콤비네이터
- 투자 범위 : Seed ~ Pre A 단계

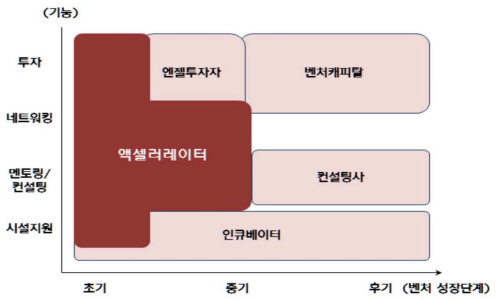
벤처캐피털

□ 상장 중소기업에게 보통주, 우선주, 상환전환우선주 등 주식의 형태로 투자하고, 각종 지원을 통해 기업 가치를 상승시킨 후 지분 매각을 통해 수익을 얻는 투자회사 (구경철, 2008)

□ 창업 초기이면서 기술력과 장래성을 가졌으나 경영 기반이 약해 큰 리스크(Risk)를 가진 벤처기업에 자금을 공급하고 경영 전반에 대해 지원하는 기업(김대민, 2001)

- 투자 범위 : Series A ~ Series C 단계

[기능과 지원 범위]



구분	금액
시드투자	~3억
프리-A	5~15억원
시리즈 A	20억~30억
시리즈 B	50억~100억
시리즈C~E	100억~수천억
IPO	주식시장 상장 (Initial PublicOffering)
M&A	기업인수합병 (Merger &Acquistion)

※ 자료 : 박희원(2018), 해외 액셀러레이터 기관 사례분석 및 시사점

Part 2 이론적 배경 (투자자의 스타트업 투자 결정요인에 관한 선행 연구)

□ 스타트업이 지속 성장을 하기 위한 필요 요인은 창업자 역량, 혁신 아이디어, 기술개발 전략 등 다양하지만 그 중에서도 가장 중요한 요인으로 기업의 확장과 발전의 기회를 제공하는 자금조달 능력 (이현호 외, 2017)

- 자금조달 역량이 경영성과에 미치는 영향이 통계적으로 유의하고 기업의 전략관리에 대한 역량이 경영성과에 미치는 영향에 있어 자금조달 역량의 완전 매개효과가 존재한다는 것을 통계적으로 검증 (손종욱 외, 2021)

□ 투자사들은 창업자 역량, 제품 및 서비스, 시장 규모, 재무적 특성, 수익률 등과 같은 '투자의사 결정 요인'에 대해 각자 다른 우선순위를 가지고 투자를 결정 (한주형. 황보윤, 2020; 구중회 등, 2019)

□ 최근 ESG 경영요소 또한 투자의사결정에 영향을 주며, 다양한 요인 중 잠재적 수익률 다음으로 스타트업의 ESG 경영요소가 투자결정의 우선순위 요인으로 투자의사결정에 큰 영향 (이태범 , 2022)

- 기업의 ESG경영은 내부적으로 경영자의 효율적인 운영을 유도하며, 외부적으로는 투자자들의 신뢰도를 높임으로써 자본에 대한 접근성을 높여 기업의 투자유치 효율성에 있어 긍정적 영향을 끼친다는 사실을 확인 (반혜정 외, 2022)

⇒ 투자자들의 투자 결정요인 간 우선순위에는 각 투자 주체별로 차이가 있으나 최근 들어 전통적인 투자 결정요인(재무적)과 함께 중요한 요인으로 ESG 경영요소(비재무적)를 인식하고 있음

Part 2 이론적 배경 (ESG 경영요소에 관한 선행 연구)

□ 기업이나 투자자가 환경, 사회, 지배구조 측면에서 지속 가능성을 고려하는 것을 의미하며, 3가지 구성요소는 다양한 하위 요인들로 구성(Ben-Amar et al., 2017)

□ ESG 경영 역량이 높을수록 조직의 효율성과 경영성과 및 기업가치의 향상을 이끌고(Gillan et al, 2010), 기업 이미지 개선과 브랜드 가치제고를 통해 장기적으로 지속가능한 성장 가능

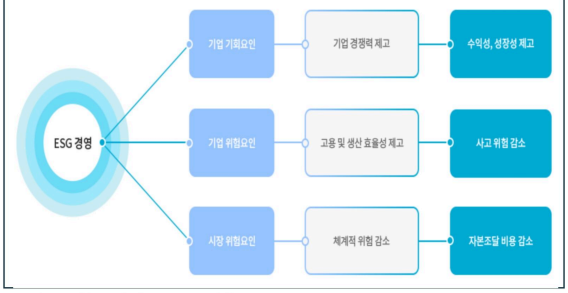
- ESG 경영 정보를 바탕으로 한 투자는 주요한 투자 전략 가운데 하나이며, ESG 경영평가가 높은 기업은 통계적으로 높은 투자 수익을 기대할 수 있는 만큼(한국ESG기준원, 2021), 투자자들은 ESG경영을 통한 수익창출 기대를 투자 포트폴리오에 반영할 가능성이 높음
- 과거에는 기업을 평가함에 있어서 제품 및 서비스를 제공하고 이익을 얻는 '재무적'인 정량 지표가 기준이었으나 글로벌 기후 위기 등 기업이 사회에 미치는 영향이 증가하여 '비재무적'인 정성 지표가 기업 가치평가에 있어 더욱 중요하다는 인식 증가

[ESG 구성 요소]

환경 (Environment)	사회 (Social)	지배구조 (Governance)
기후변화 및 탄소 배출	데이터 보호 및 프라이버시	이사회 및 감사위원회 구성
환경오염 및 환경규제	인권, 성별 평등 및 다양성	뇌물 및 반부패
생태계 및 생물 다양성	지역사회 관계	기업윤리

※ 자료 : 한국거래소 ESG 포털

[ESG 경영을 통한 가치 제고]



※ 자료 : 한국거래소 ESG 포털

참고자료 한국ESG기준원 ESG 평가체계

□ 한국ESG기준원은 환경과 사회책임 그리고 지배구조의 올바른 경영전략을 제시한 ESG 기준을 2010년부터 발표, 일반상장사를 대상으로 2011년부터 ESG 평가 진행

대분류	중분류	소분류
환경 (Environment)	리더십과 거버넌스	거버넌스 / 전략 및 목표 / 환경경영 내재화
	위험관리	위험관리
	운영 및 성과	기후변화 / 오염물질·화학물질 / 친환경 제품 및 서비스 등
	이해관계자 소통	이해관계자 대응 / 환경정보 공개
사회 (Social)	리더십과 거버넌스	전략과 방침
	노동관행	공정하고 차별없는 고용 / 일과 생활의 균형 지원 등
	직장 내 안전보건	안전보건 거버넌스 / 안전보건 위험 관리
	인권	인권경영 거버넌스 / 인권 위험 관리
	공정운영 관행	공정운영 거버넌스 / 공정운영 위험관리 등
	지속가능한 소비	소비자 권익 보호 거버넌스 / 소비자와의 소통 등
	정보보호 및 개인정보보호	정보보호 및 개인정보보호 거버넌스 / 정보보호 투자 등
	지역사회 참여 및 개발	지역사회 참여 거버넌스 / 지역사회 위험관리 등
	이해관계자 소통	사회책임경영 정보 공개
지배구조 (Governance)	이사회리더십	이사회 독립성 / 이사회 다양성 및 전문성 등
	주주권 보호	주주의 권리 / 주주 환원 / 소유구조
	감사	감사기구 구성 / 감사기구 운영 등
	이해관계자 소통	정보공개

Part 3 질적 연구 (연구 방법)

□ 본 연구의 목적은 창업자 및 투자자가 가지고 있는 ESG 경영에 대한 인식 수준과 ESG 경영요소 도입에 대한 견해를 탐구하는 것이며, 창업자 및 투자자 입장의 의견을 수집하여ESG 경영요소 도입 우선순위에 영향을 주는 요인이 무엇인지를 질적(인터뷰)으로 탐구

[인터뷰 대상자 인구 통계적 특성]

인터뷰 대상자	성별	국적	연령	직업	업력
창업자 1	남	한국	30대	창업자	5년
투자자 1	남	한국	40대	투자자	7년

[창업자 대상 인터뷰 문항]

구분	질문
(1) ESG 경영요소 도입과 관련한 인식	ESG 경영에 대해 들어본 적이 있습니까?
	귀하의 기업에서 ESG 경영 요소를 도입한다면 그 이유는 무엇인가요?
	ESG 요소(환경, 사회, 지배구조) 중에서 귀하의 기업에 가장 중요한 요소는 무엇이며, 그 이유는 무엇인가요?
	창업 초기 단계에서 ESG 경영을 고려하는 것이 어렵거나 부담스러웠던 경험이 있나요? 있다면 이유는 무엇인가요?
(2) ESG 도입의 우선순위 및 결정 과정	ESG 요소를 도입할 때 어떤 기준으로 우선순위를 결정하나요?
	투자자나 고객의 요구가 ESG 도입 결정에 영향을 미친 경험이 있나요?
(3) 투자자와의 관계 및 ESG 기대치	투자자가 귀하의 기업에 ESG 관련 요구를 한 적이 있나요? 그렇다면 어떤 요구였나요?
	ESG 요소가 기업 가치 평가나 투자 유치에 긍정적인 영향을 준다고 생각하시나요?

[투자자 대상 인터뷰 문항]

구분	질문
(1) 투자 의사결정에서 ESG 고려 수준	ESG 경영에 대해 들어본적이 있습니까?
	투자 대상 기업을 평가할 때 ESG 요소(환경, 사회, 지배구조) 중 가장 중시하는 요소는 무엇이며, 그 이유는 무엇인가요?
	ESG 요소가 미흡한 기업이라도 재무적 성과가 우수하다면 투자하시겠습니까?
(2) ESG 요소 도입에 대한 요구 및 우선순위	투자 대상 기업이 ESG 요소를 도입하는 과정에서 투자자로서 요구하는 사항이 있나요?
	특정 산업(예: 제조업, IT, 바이오 등)에 따라 ESG 요소에 대한 기대가 달라지나요?
	ESG 경영이 장기적인 투자 수익률에 영향을 미친다고 보시나요? 그렇다면 이유는 무엇인가?
(3) 창업자와의 의견 차이 및 과제	투자자와 창업자 간 ESG 도입 우선순위에 대해 가장 큰 견해 차이는 무엇이라고 생각하시나요?
	ESG 경영을 강화하기 위해 투자자로서 창업자들에게 제공할 수 있는 조언이 있다면 무엇인가요?

Part 3 질적 연구 (연구 결과)

□ 창업자 : 사회 요소를 중요하게 생각, 현재는 도입 여력이 부족해 ESG 도입을 기업 성장 단계에 따라 전략적으로 접근해야 한다고 판단
투자자 : 지배구조 요소를 최우선으로 인식, 창업 기업이 장기적 기업 가치를 높이려면 초기부터 ESG를 고려해야 한다고 인식

⇒ 창업자와 투자자 모두 투자유치 및 기업 성장 과정에서 ESG의 중요성이 커지고 있으며, 스타트업도 점진적으로 ESG 체계를 도입해야 경쟁력을 확보 할 수 있다고 보고 있음

[인터뷰 결과 종합 요약]

구분	창업자	투자자
ESG 인식	**ESG가 중요하다는 것은 알지만, 현재 도입할 여력은 부족**	**ESG는 필수 요소이며, 초기부터 준비해야 한다고 인식**
ESG 요소 중시 순위	**사회 요소 중시 (브랜드 이미지, 소비자 신뢰 등)**	**지배구조 최우선 (이사회 운영 등), 산업별로 사회 및 환경 요소도 고려**
ESG 도입 필요성	**브랜드 파워 강화 및 투자 유치 신뢰 확보**	**장기적 리스크 관리 및 기업 가치 상승**
산업별 ESG 기준	IT 기업으로서 데이터 보호 및 개인정보 관리 중시	제조업: 환경 / IT: 데이터 보안 / 바이오: 윤리성
ESG 도입 시 우선순위	기업 성장 단계 및 정책에 따라 점진적 도입, 투자자가 요구하는 요소 우선 적용 가능	지배구조 관련 요구 (이사회 운영, 내부 규정, 의사결정 절차 검토)
장기 투자 수익률	ESG를 실천한 기업은 소비자와 투자자로부터 긍정적 평가를 받을 것	ESG 도입 기업이 글로벌 시장 진출, IPO 및 M&A에서 유리
도입의 어려움	**인적·물적 자원 부족으로 초기 단계에서 ESG를 적용 어려움**	**초기 창업자가 비용 부담을 이유로 ESG 도입을 기피하는 것이 문제**
투자자의 ESG 요구	분기별 이사회 운영 및 기업 운영 관련 자료 요구를 받고 있음	기업 운영의 투명성에 대한 요구 (이사회 운영, 주주총회, 직원 복지 등)
ESG 도입에 대한 조언	**ESG 경영을 준비하지 않으면 향후 투자 유치 및 정부 지원사업 참여가 어려워질 것**	**초기부터 ESG 체계를 준비, 정부·창업지원 기관의 ESG 지원 활용**

Part 4 양적 연구 (연구 문제 및 요인 선정)

□ 본 연구는 그간의 선행 연구를 기반으로 ESG 경영의 상위 요인과 하위 요인을 도출하여 범주화하고, 디지털 기업 창업자 및 투자자를 대상으로 ESG 경영요소 도입의 우선순위를 평가
다음으로 창업자와 투자자의 ESG 경영요소 도입 우선순위 인식의 차이점을 확인하고 비교·분석 수행

[연구 문제]

연구문제 1	디지털 스타트업 관점에서 ESG 경영요소 도입 시 우선순위는 어떻게 되는가?
연구문제 2	투자자 관점에서 디지털 스타트업의 ESG 경영요소 도입 시 우선순위는 어떻게 되는가?
연구문제 3	디지털 스타트업과 투자자 간 ESG 경영요소 도입의 우선순위의 공통점과 차이점은 무엇인가?

[변수 설정]

상위 요인	하위 요인	항목 설명
환경	-	자연환경 보호 및 기후 변화, 탄소중립, 지속가능자원 활용 등
사회	-	인권보호, 노동관행 개선, 지역사회 발전 등
지배구조	-	주주권리 보호, 의사결정구조 및 윤리적 경영 등
환경 (Environment)	최고경영자의 실천의지	최고경영자 실천의지 표명, 환경정보 공개 등
	환경경영 전략 및 방침	기업의 경영전략 및 방침과 통합 수립 여부 등
	환경친화적 생산	제품 및 서비스 전 과정에서 예방차원의 환경관리 등
	환경보호 활동	국내외 환경관련 프로그램의 자발적 참여 등
사회 (Social)	직장 내 안전 및 보건	근로자의 안전보건 정책 수립 유무 등
	직장 내 기본권 보호	공정한 고용, 일과 생활 균형 지원, 차별금지 정책 등
	기업경영 부패 방지	부패방지 전담조직 유무, 윤리교육 실시 여부 등
	소비자 개인정보 보호	소비자 개인정보 관리 정책 수립 등
지배구조 (Governance)	주주의 권리 보호	소유자로서의 권리, 주주의 권리의 최대 보장 등
	이사회 기능 활성화	기업 의사결정 기능 및 감독 기능 수행 등
	감사 기능 강화	외부 감사기구 구성, 감사기구 독립적 운영 등
	기업 공시 투명성	의사결정에 중요한 영향을 미치는 사항에 대한 공시 등

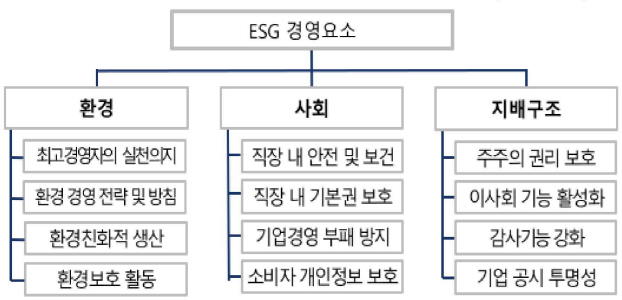
※ 자료 : 남정민·박준혁·이소정(2022)의 연구를 바탕으로 본 연구에 맞게 연구자 재정리

Part 4 양적 연구 (연구 방법)

□ 인간이 복잡한 의사결정을 할 때 계층 구조화 및 상대적 중요도 추정, 논리적 일관성 확보 등과 같은 프로세스를 거친다는 점에서 착안된 분석적 계층화법(AHP, Analytic Hierachy Process)을 활용하여 설문조사 진행

창업 7년 이내 디지털 스타트업 창업자 15명, 벤처캐피털 및 액셀러레이터 등 전문 투자자 15명을 대상으로 2025년 4월 2일부터 5월 20일까지 7주간에 걸쳐 설문조사를 실시

[ESG 경영요소 주요 요인 계층 구조]



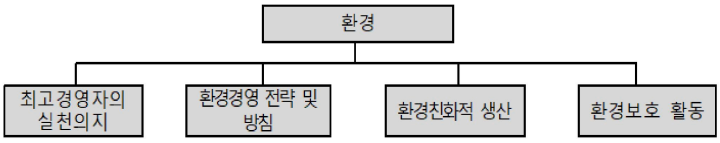
[조사 대상자의 인구통계적 특성]

항목	구분	창업자		투자자	
		빈도	비율(%)	빈도	비율(%)
성별	남	12	80.0	11	73.3
	여	3	20.0	4	26.6
연령	30대	4	26.6	5	33.3
	40대	11	73.3	10	66.6
학력	대학 졸업	8	53.3	7	46.6
	대학원 이상	7	46.6	8	53.3
업력	3년 미만	3	20.0	1	6.6
	3년 이상~5년 미만	8	53.3	3	20.0
	5년 이상~7년 미만	4	26.6	6	40.0
	7년 이상	-	-	5	33.3
유형	액셀러레이터	-	-	6	40.0
	벤처캐피털			9	60.0

참고자료

AHP분석 쌍대 비교 예시

[쌍대 비교 예시]



항목	항목설명
최고경영자의 실천의지	최고경영자 실천의지 표명, 환경정보 공개 등
환경경영 전략 및 방침	기업의 경영전략 및 방침과 통합 수립 여부 등
환경친화적 생산	제품 및 서비스 전 과정에서 예방차원의 환경관리 등
환경보호 활동	국내외 환경관련 프로그램의 자발적 참여 등

항목	중요←				동등				→ 중요	항목
	5	4	3	2	1	2	3	4	5	
최고경영자의 실천의지										환경경영 전략 및 방침
최고경영자의 실천의지										환경친화적 생산
최고경영자의 실천의지										환경보호 활동
환경경영 전략 및 방침										환경친화적 생산
환경경영 전략 및 방침										환경보호 활동
환경친화적 생산										환경보호 활동

[쌍대 비교의 상대적 중요성에 대한 척도]

중요도	정의	설명
1	동등함	두 요인의 중요도가 동일하거나 비슷함
2	약간 중요함	두 요인을 비교할 때 어떤 요인이 약간 중요함
3	중요함	두 요인을 비교할 때 어떤 요인이 확실히 중요함
4	매우 중요함	두 요인을 비교할 때 어떤 요인이 매우 중요함
5	절대 중요함	두 요인을 비교할 때 어떤 요인이 절대 중요함

Part 4

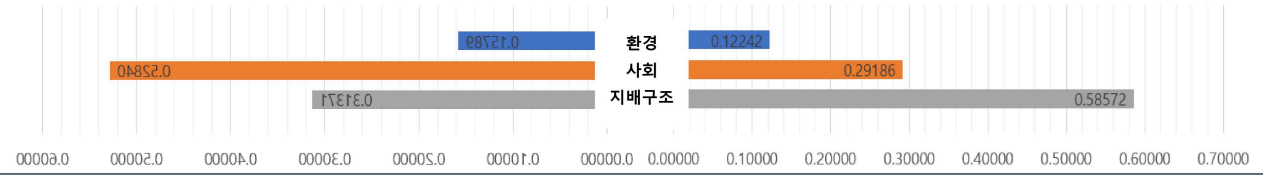
양적 연구 (연구 결과-상위계층 요인)

[창업자 상위계층 요인 가중치]

선택기준	환경	**사회**	지배구조	계
가중치	0.15789	**0.52840**	0.31371	1
순위	3	**1**	2	
CR = 0.00481				

[투자자 상위계층 요인 가중치]

선택기준	환경	사회	**지배구조**	계
가중치	0.12242	0.29186	**0.58572**	1
순위	3	2	**1**	
CR = 0.02493				



- 1순위 : 창업자들은 소비자의 개인정보, 데이터 윤리, 노동 환경 개선 등의 사회를 1순위로 생각, 투자자는 주주로서 권리와 이사회 운영 등 투자자의 권리를 위한 지배구조의 우선적인 도입을 고려하고 있는 것을 확인
- 2순위 : 창업자는 투자유치 시 필요한 주주 권리, 윤리 경영과 같은 지배구조를 차순위로 고려, 투자자는 디지털 창업 기업의 지속 성장 및 매출에 영향을 주는 기업 이미지와 연관된 소비자 보호, 노동 환경 개선 등의 사회를 고려
- 3순위 : 디지털 창업 기업 특성상 탄소중립, 기후 위기와 같은 환경 요인과 직접적 영향을 주고 받는 구조로 인식하지 않아 창업자와 투자자 모두 환경을 가장 후 순위로 고려

Part 4 양적 연구 (연구 결과-상위계층 요인)

[창업자 상위계층 요인 쌍대비교 분석]

상위 요인	환경	사회	지배구조	가중치
환경	1.000	0.321	0.467	0.15789
사회	**3.112**	1.000	**1.817**	0.52840
지배구조	**2.143**	0.550	1.000	0.31371

창업자의 상위 요인 쌍대 비교 분석에 따르면 ESG 경영요소 도입 우선순위는 사회의 중요도가 각각 환경, 지배구조의 3.112배, 1.817배, 지배구조의 중요도가 환경의 2.143배

창업자들은 ESG 경영요소 도입 시 사회, 지배구조, 환경 순으로 고려하는 것으로 확인

[투자자 상위계층 요인 쌍대비교 분석]

상위 요인	환경	사회	지배구조	가중치
환경	1.000	0.352	0.244	0.12242
사회	**2.837**	1.000	0.417	0.29186
지배구조	**4.090**	**2.398**	1.000	0.58572

투자자의 상위 요인 쌍대 비교 분석에 따르면 ESG 경영요소 도입 우선순위는 지배구조의 중요도가 각각 환경, 사회의 4.090배, 2.398배, 사회의 중요도가 환경의 2.837배

투자자들은 ESG 경영요소 도입 시 지배구조, 사회, 환경 순으로 고려하는 것으로 확인

Part 4 양적 연구 (연구 결과-하위계층 요인)

<창업자 하위 요인>

[환경에 대한 하위계층 요인 가중치]

선택 기준	**최고경영자의 실천의지**	**환경경영 전략 및 방침**	환경친화적 생산	환경보호 활동
가중치	**0.43122**	**0.28001**	0.18279	0.10598
순위	**1**	**2**	3	4
CR = 0.01034				

[사회에 대한 하위계층 요인 가중치]

선택 기준	직장 내 안전 및 보건	**직장 내 기본권 보호**	기업경영 부패 방지	**소비자 개인정보 보호**
가중치	0.11438	**0.25819**	0.18249	**0.44495**
순위	4	**2**	3	**1**
CR = 0.01054				

[지배구조에 대한 하위계층 요인 가중치]

선택 기준	주주의 권리 보호	**이사회 기능 활성화**	감사 기능 강화	**기업 공시 투명성**
가중치	0.20318	**0.29945**	0.15411	**0.34325**
순위	3	**2**	4	**1**
CR = 0.00045				

<투자자 하위 요인>

[환경에 대한 하위계층 요인 가중치]

선택 기준	**최고경영자의 실천의지**	**환경경영 전략 및 방침**	환경친화적 생산	환경보호 활동
가중치	**0.31045**	**0.32461**	0.22416	0.14078
순위	**2**	**1**	3	4
CR = 0.00142				

[사회에 대한 하위계층 요인 가중치]

선택 기준	직장 내 안전 및 보건	**직장 내 기본권 보호**	기업경영 부패 방지	**소비자 개인정보 보호**
가중치	0.12021	0.17473	**0.35875**	**0.34631**
순위	4	3	**1**	**2**
CR = 0.00764				

[지배구조에 대한 하위계층 요인 가중치]

선택 기준	주주의 권리 보호	**이사회 기능 활성화**	감사 기능 강화	**기업 공시 투명성**
가중치	**0.43750**	**0.23844**	0.20467	0.11938
순위	**1**	**2**	3	4
CR = 0.00327				

Part 4 양적 연구 (연구 결과-하위계층 요인)

<창업자 및 투자자 하위계층 요인 중요도 비교>

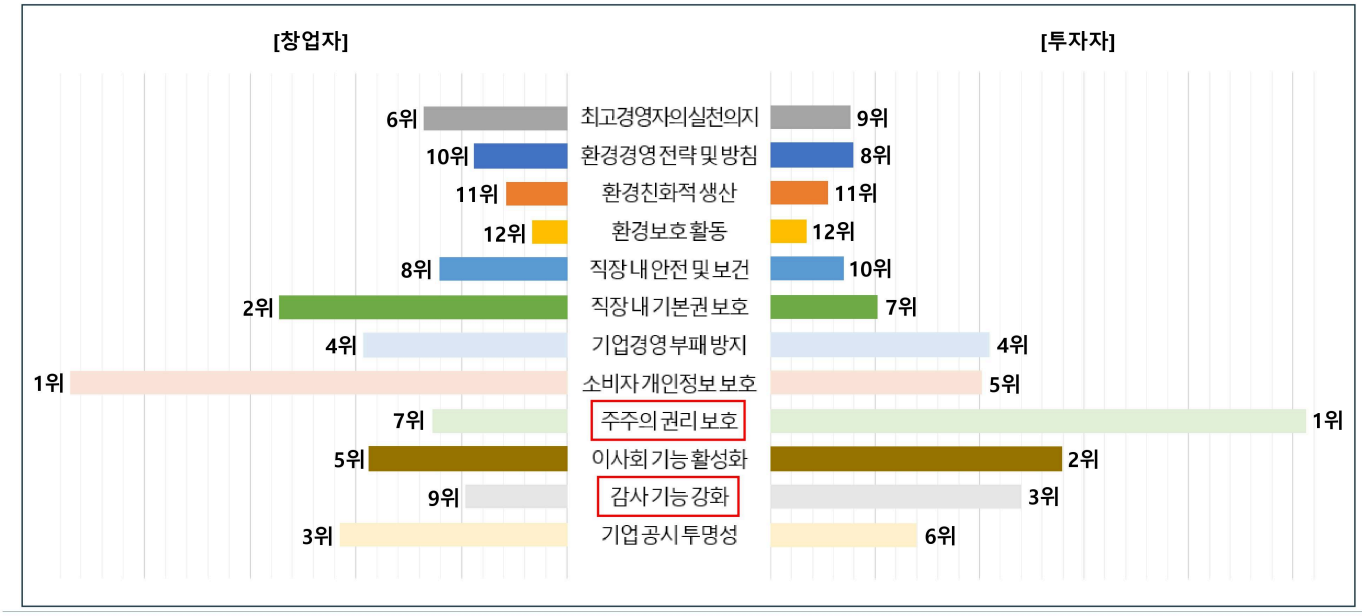
창업자		하위계층 요인	투자자	
가중치	순위		순위	가중치
0.43122	**1**	최고경영자의 실천의지	**2**	0.31045
0.28001	**2**	환경경영 전략 및 방침	**1**	0.32461
0.18279	3	환경친화적 생산	3	0.22416
0.10598	4	환경보호 활동	4	0.14078
0.11438	4	직장 내 안전 및 보건	4	0.12021
0.25819	**2**	직장 내 기본권 보호	3	0.17473
0.18249	3	기업경영 부패 방지	**1**	0.35875
0.44495	**1**	소비자 개인정보 보호	**2**	0.34631
0.20318	3	주주의 권리 보호	**1**	0.43750
0.29945	**2**	이사회 기능 활성화	**2**	0.23844
0.15411	4	감사 기능 강화	3	0.20467
0.34325	**1**	기업 공시 투명성	4	0.11938

Part 4 양적 연구 (연구 결과-하위계층 요인 전체 가중치 종합 비교)

창업자				투자자			
순위	전체 가중치	하위계층 요인	상위계층 요인	상위계층 요인	하위계층 요인	전체 가중치	**순위**
6	0.06809	최고경영자 실천의지 (0.43122)	환경 (0.15789)	환경 (0.12242)	최고경영자 실천의지 (0.31045)	0.03801	9
10	0.04421	환경경영 전략 및 방침 (0.28001)			환경경영 전략 및 방침 (0.32461)	0.03974	8
11	0.02886	환경친화적 생산 (0.18279)			환경친화적 생산 (0.22416)	0.02744	11
12	0.01673	환경보호 활동 (0.10598)			환경보호 활동 (0.14078)	0.01723	12
8	0.06044	직장 내 안전 및 보건 (0.11438)	사회 (0.52840)	사회 (0.29186)	직장 내 안전 및 보건 (0.12021)	0.03509	10
2	**0.13643**	**직장 내 기본권 보호 (0.25819)**			직장 내 기본권 보호 (0.17473)	0.05100	7
4	**0.09643**	**기업경영 부패 방지 (0.18249)**			**기업경영 부패 방지 (0.35875)**	**0.10470**	**4**
1	**0.23511**	**소비자 개인정보 보호 (0.44495)**			**소비자 개인정보 보호 (0.34631)**	**0.10107**	**5**
7	0.06374	주주의 권리 보호 (0.20318)	지배 구조 (0.31371)	지배 구조 (0.58572)	**주주의 권리 보호 (0.43750)**	**0.25625**	**1**
5	**0.09394**	**이사회 기능 활성화 (0.29945)**			**이사회 기능 활성화 (0.23844)**	**0.13966**	**2**
9	0.04835	감사 기능 강화 (0.15411)			**감사 기능 강화 (0.20467)**	**0.11988**	**3**
3	**0.10768**	**기업 공시 투명성 (0.34325)**			기업 공시 투명성 (0.11938)	0.06993	6

Part 4 양적 연구 (연구 결과-하위계층 요인 전체 가중치 종합 비교)

<하위계층 요인 전체 가중치 종합 비교 차트>



Part 5 결 론 (연구 결과 요약)

창업자	투자자
▶ 창업 초기 ESG 도입은 어려우나 기업 성장에 따라 도입은 필수	▶ 스타트업의 지속성장을 위해 창업 초기부터 ESG 도입 필요
▶ '사회'를 가장 중시, 창업 초기 단계 직접적 이해관계자인 소비자와 내부 구성원에 대한 책임 중시	▶ '지배구조'를 가장 중시, 투자의 안정성과 리스크 관리, 경영 투명성 확보를 중시
▶ 특히, 소비자 개인정보 보호와 직장 내 기본권 보호 등 신뢰 기반의 고객 관계와 건강한 조직문화 조성을 중요하게 인식	▶ 특히, 주주의 권리 보호 와 이사회 기능 활성화, 감사 기능 강화 등 자본 제공자로서 주주 권리 확보, 내부 감독 체계를 중요하게 인식
▶ '지배구조' 에서는 기업 공시 투명성을 우선 요인으로 인식, 기업 제도 개선보다 외부 이해관계자와의 신뢰 형성을 더욱 중시	▶ '사회' 에서는 기업경영 부패 방지와 소비자 개인정보 보호를 중요하게 평가, 이는 사회적 신뢰와 브랜드 이미지가 기업 가치와 직결된다는 판단에 기반

▶ 창업자가 사회적 책임 중심의 ESG 경영요소를 중시하는 반면, 투자자는 기업 내부 제도 중심의 ESG 경영요소를 중시

▶ 창업자와 투자자 간 ESG 인식의 관점 차이를 이해하고, 상호 간의 의견 조율을 위한 소통 전략이 필요함을 시사

Part 5 결 론 (연구 시사점)

학술적 의의

- ESG 경영이 주로 대기업 중심으로 논의되던 상황에서 디지털 초기 창업 기업과 민간 투자자의 인식 차이를 정량·정성적으로 분석한 점에서 의의

  또한 ESG 경영요소의 하위계층 요인들의 우선순위를 창업 생태계 핵심 주체인 창업자와 투자자를 대상으로 실증적 비교 분석을 수행

- ESG 경영요소 인식 차이를 대리인 이론의 틀에서 해석할 수 있는 실증적 근거를 제시

  대리인(경영자)의 기회주의적 행동을 예방할 수 있는 통제 기제를 ESG 도입의 핵심 요소로 인식하고 있음을 확인

- 질적 연구(심층 인터뷰)와 양적 연구(AHP 분석)를 결합한 혼합 방법론을 활용함으로써 단일 방법론의 한계를 극복했다는 점에서 의의

실무적 의의

- 디지털 창업 기업이 기술개발과 생존 중심의 전략에 매몰되어 지배구조 요소의 중요성을 간과하는 '기술 근시(technological myopia)' 현상을 드러내고 있다는 점에서 중요한 함의

  디지털 창업 기업은 창업 초기부터 기술개발 전략과 병행하여 ESG 내재화 전략을 장기적 관점에서 설계하고 실천할 필요

  투자자가 중시하는 지배구조 관련 ESG 요소들과 ESG 인식 간 우선순위 차이를 인지하고, 이를 투자유치에 유리한 방향으로 사전에 조율하는 전략적 노력이 요구

- ESG 기반 투자 심사 기준을 설계하는 데 실질적 활용 가능성 제시

  투자자들은 본 연구 결과를 통해 창업자가 어떤 ESG 요소를 우선으로 고려하는지, 자원의 제약 속에서 ESG를 어떻게 현실화하고 있는지에 대한 실질적 이해를 바탕으로, 투자 심사 체크리스트나 평가 프레임워크를 보다 정교화 할 필요성

Part 5 결 론 (한계 및 향후 연구 방향)

한계

질적 연구를 위한 인터뷰가 매우 제한적이었으며 정량적 설문조사 역시 규모가 각각 15명으로 제한적으로 일반화 제약

디지털 스타트업에 국한되어 있어 제조, 바이오 등 타 산업군에서의 ESG 경영요소의 중요도 인식이 다를 수 있다는 점

향후 연구 방향

더 많은 수의 창업자 및 투자자를 대상으로 한 확장된 분석 필요, 다양한 배경과 특성을 반영한 세분화된 비교가 요구
(응답자의 성별, 연령, 기업의 성장 단계, 투자유치 규모 등 인구 통계적 및 조직적 변수에 따른 교차 분석)

제조업, 헬스케어, 바이오, 친환경 산업 등 다양한 산업군으로 확장하는 것이 필요
(산업별로 창업 기업이 직면하는 환경적 영향, 사회적 책임 수준, 법적·제도적 규제의 강도, 이해관계자의 구성과 요구가 상이하므로 ESG 요소의 상대적 중요도와 도입 방식 또한 산업 특성에 따라 달라질 가능성)

↓

향후 연구에서는 표본 수 확대를 통한 일반화 가능성 제고와 산업군 확장을 통한 다양성 확보라는 두 축을 중심으로 ESG 인식 및 도입 전략에 대해 보다 실증적이고 입체적인 논의가 이루어질 필요

	Memo	

Track
4-2

온라인 플랫폼의 위험관리 전략

좌장: 이중엽 소프트웨어정책연구소 책임연구원

소셜미디어 플랫폼의 악성콘텐츠 대응 전략

임희주(서원대학교)

AI 모델의 프라이버시-신용 위험 예측 정확도 상충 관계 분석을 통한 소비자 신용 위험 관리

양은경(연세대학교)

From YouTube to Reddit: Cross-Platform Topic Propagation in World of Warcraft

르키아(순천향대학교)
유여완(순천향대학교)
최재원(순천향대학교)

Behind the Stars: A Text Mining Study of Negative Reviews and Seller Strategies

김엘레나(순천향대학교)
최재원(순천향대학교)

소셜미디어 플랫폼의 악성콘텐츠 대응 전략

임희주
서원대학교 조교수

본 연구는 악성콘텐츠가 증가하는 소셜미디어 환경에서 소셜미디어 플랫폼의 안전투자와 광고비 결정에 대한 최적 전략을 분석한다. 선행 연구가 콘텐츠 조정과 광고 전략을 독립적으로 다루었다는 한계를 극복하고, 비대칭 복점 경쟁 상황 하에 있는 소셜미디어 플랫폼에게 있어, 이 두 요소에 대한 전략을 통합적으로 규명하는 것을 목적으로 한다.

양면시장 플랫폼 이론에 기초한 해석적 연구 모형을 구축하여 두 개의 비대칭 플랫폼이 사용자와 광고주를 동시에 유치하기 위해 경쟁하는 상황을 분석한다. 분석 결과는, 악성콘텐츠 증가 시 우위 플랫폼은 안전투자와 광고가격을 함께 증가시키는 반면, 열위 플랫폼은 광고비를 일관되게 인하하고 안전투자는 우위 플랫폼과의 경쟁 격차(플랫폼 고유 가치의 차이)에 따라 조건부로 결정됨이 나타났다. 즉, 두 플랫폼과의 경쟁 격차가 임계값보다 작으면 열위 플랫폼도 안전투자를 증가시키지만 임계값을 초과하면 비용 절감을 위해 안전투자를 감소시킨다. 또한 악성콘텐츠 증가 시 후생 변화를 분석한 결과, 고객의 총효용과 우위 플랫폼 이윤은 각각의 경쟁 격차에 따라 증가 또는 감소할 수 있으나 열위 플랫폼 이윤은 항상 감소하고, 플랫폼의 방어적 광고비 전략에 기인해 광고주의 총이윤은 항상 증가하는 역설적인 결과를 도출하였다.

본 연구는 양면시장 프레임워크 내에서 콘텐츠 조정(안전투자)과 광고가격을 통합적으로 고려한 연구모형을 수립해 분석함으로써, 우위 플랫폼은 안전품질 차별화를 통한 프리미엄 전략을, 열위 플랫폼은 경쟁 격차 크기에 따라 조건부 전략을 선택해야 함을 확인하였다는 점에서 실무적 공헌을 찾을 수 있다. 또한, 양면시장 프레임워크 내에서 콘텐츠 조정 및 광고비 전략을 동시에 고려한 연구모형을 수립함으로써, 소셜미디어 플랫폼의 핵심 비즈니스 요소 간 전략적 상호의존성을 규명하였다는 점에서 학술적 기여를 설명할 수 있다.

KEYWORDS 소셜미디어, 양면시장, 플랫폼 경쟁, 광고 전략, 안전투자

AI 모델의 프라이버시-신용 위험 예측 정확도 상충 관계 분석을 통한 소비자 신용 위험 관리

양은경
연세대학교 정보대학원 정보보호 학과

AI와 인간의 교류는 점점 가속화될 것이다. 하지만 모방이나 통계적 추론은 가능해도 AI 입장에서 인간에 대해 제대로 “이해”하기 어려운 특성들이 있을 것으디지털 금융의 확산은 인공지능(AI)을 소비자 신용 리스크 관리의 핵심 기술로 자리 잡게 하지만, 대규모 개인신용정보 활용은 프라이버시 침해 위험을 동반한다. 이에 따라 금융기관은 예측 성능을 유지하면서도 프라이버시 보호 요구를 충족해야 하는 과제에 직면하고 있다. 본 연구는 차등 프라이버시(Differential Privacy, DP)를 적용한 신용 리스크 예측 모델을 대상으로 프라이버시–정확도 상충 관계(privacy–accuracy trade-off)를 정량적으로 분석하는 연구 프레임워크를 제시한다.로 보인다. 대표적으로, 왜 인간은 행복과 같은 감정적 경험을 그토록 중시할까? 강의를 통해 최근의 진화심리학 관점으로 재해석하는 행복(감정)의 기능과 존재 이유를 소개한다. AI와 더불어 사는 시대에 여러 기술적, 철학적 함의점이 담긴 질문이다.

본 연구는 공개 벤치마크 데이터셋인 UCI Default of Credit Card Clients를 활용하여 Gradient Boosting 계열 세 모델(XGBoost, LightGBM, CatBoost)을 비교 분석한다. Laplace 메커니즘 기반 차등 프라이버시 환경에서 프라이버시 예산(ε)을 [∞, 10, 5, 1, 0.5, 0.1]로 설정하여 모델 성능 변화를 AUC와 AUPRC로 평가하고, t-test, ANOVA, Tukey HSD를 통해 ε 변화에 따른 성능 차이의 통계적 유의성을 확인한다. 분석 결과, ε이 감소할수록 모델 성능은 비선형적으로 하락하며, ε≈1 구간에서 급격한 성능 저하가 발생한다. 또한 동일한 프라이버시 강도에서도 모델별 성능 저하 폭이 상이한 프라이버시 민감도(model-specific sensitivity)가 존재한다. LightGBM은 상대적으로 안정적인 반면, CatBoost는 상대적으로 높은 민감도를 나타내어 프라이버시 예산 설정 시 모델 구조적 특성의 고려가 필요함을 시사한다.

본 연구는 기존 연구가 단일 모델 중심으로 프라이버시 비용을 논의해온 범위를 넘어, 모델별 비교를 통해 프라이버시 효율성(privacy efficiency) 개념을 제시한다는 점에서 학문적 기여를 갖는다. 또한 금융기관이 목표 AUC 수준과 리스크 허용 범위에 따라 적정한 ε 값 조정을 수행할 수 있는 의사결정 근거를 제공한다. 이 연구는 프라이버시 보호와 데이터 활용이 대립적 선택이 아니라 관리·조정 가능한 상충 구조임을 보여주며, Responsible AI와 금융 데이터 거버넌스 구축을 위한 실무적 기반을 마련한다.

KEYWORDS 차등 프라이버시, 프라이버시–정확도 상충 관계, 신용 리스크 예측, Gradient Boosting, Responsible AI

From YouTube to Reddit: Cross-Platform Topic Propagation in World of Warcraft

르키아
순천향대학교 경영학과
–
유여완
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

This study investigates the impact of thumbnail-title emotional congruence on audience engagement in branded YouTube content, employing a hybrid machine-learning and statistical approach. A multimodal pipeline was designed by combining CLIP-based visual and textual embeddings, followed by 5 ML classifiers to predict the thumbnail-title congruence automatically. Among the classifiers (DT, ET, RF, XGBoost, and LGBM), LGBM achieved strong predictive performance: AUC = 0.82, AP = 0.87, on a data sample of 8082 thumbnail-title pairs labeled by human encoders. LGBM then automatically detected the probability of congruence of the full dataset (125,237 branded videos). These predicted probabilities were subsequently analyzed using a log-linear regression model, which revealed a significant positive relation between congruence and view-based engagement with $\beta = 1.44$, $p < .001$, $R^2 = 0.117$. A 0.10 increase in the predicted congruence corresponded to approximately a 15.5% rise in expected views, demonstrating that emotionally aligned thumbnails and titles enhance audience attention and video visibility, validating the effectiveness of the hybrid ML-statistical modeling framework.

KEYWORDS Thumbnail-title congruence, Multimodal ML, View-based engagement, YouTube branded videos, Hybrid ML-statistical model

Behind the Stars: A Text Mining Study of Negative Reviews and Seller Strategies

김엘레나
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

In the evolving landscape of e-commerce, negative online reviews play a crucial role in shaping consumer perceptions and influencing future purchasing decisions. This study applies text mining techniques to analyze online customer reviews, focusing on two product categories—hedonic and utilitarian—and their corresponding seller responses. By employing sentiment analysis and Latent Dirichlet Allocation (LDA) topic modeling, the research identifies key themes in negative reviews, highlighting the specific aspects of the purchasing experience that consumers emphasize when expressing dissatisfaction.

Furthermore, the study investigates seller response patterns to evaluate the effectiveness of different response strategies in mitigating the impact of negative feedback. The sentiment analysis results reveal that consumer concerns differ by product type: buyers of utilitarian products primarily underscore functional failures, whereas buyers of hedonic products tend to emphasize emotional dissatisfaction.

These findings provide actionable insights for businesses, contributing to a deeper understanding of consumer behavior in the context of negative reviews. The results can guide the development of more effective crisis management strategies and optimized response mechanisms, ultimately fostering consumer trust and encouraging repurchase intentions.

KEYWORDS E-Commerce, Negative Online Reviews, Text Mining, Sentiment Analysis, Topic Modeling, Customer Experience

	Memo	

Track
4-3

전기차관련 인식 및 구매 의사결정

좌장: 고준 전남대학교 교수

소비자 불만이 전기차 위험 인식과 회피 의도에 미치는 영향: HBM과 TTAT 관점에서

왕문준(순천향대학교)
최재원(순천향대학교)

Exploring the Expectation–Mismatch in Electric Vehicle Adoption: Insights from Negative Consumer Reviews on Reddit

정자강(순천향대학교)
왕문준(순천향대학교)
최재원(순천향대학교)

플랫폼 기반 타이어 구매의도의 결정요인과 쇼루밍 성향의 조절효과

김상준(경희대학교)
양성병(경희대학교)

소비자 불만이 전기차 위험 인식과 회피 의도에 미치는 영향: HBM과 TTAT 관점에서

순천향대학교 일반대학원 경영학과

왕문준,최재원

01.서론

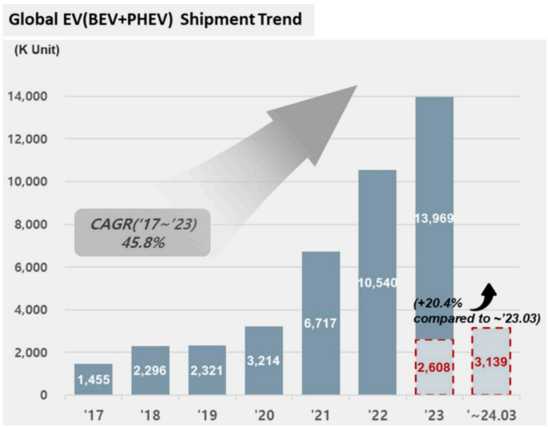
연구 배경 및 문제 인식

- 글로벌 전기차 시장은 이미 고속 성장 단계에서 안정적이고 구조적인 전환 단계로 진입함.
- 소비자들은 기대와 실제 사용 경험 사이에 뚜렷한 차이가 있음. 예를 들어, 충전 인프라 부족, 주행 거리 불안, 배터리 안전 우려 등의 문제로 인해 온라인 댓글에서 많은 부정적인 반응이 나타나고 있음.
- 이러한 소비자 불만은 사용 과정에서의 실제 문제점을 반영할 뿐만 아니라, 기업의 기술 및 서비스 혁신과 정부 정책 조정에 중요한 단서를 제공함.
- 따라서 본 연구는 이전 텍스트 마이닝 연구를 기반으로 설문조사를 통해 소비자 불만-위험 인식-행동 의도 간의 구조적 관계를 추가로 검증함.

앞선 연구(Wang et al., 2025)는 Reddit 데이터를 분석하여 소비자 불만의 세 가지 핵심 요인(기술적, 사용성, 신뢰성)을 도출하였음. 본 연구는 이를 기반으로 이론 모형을 수립하고 설문을 통해 구조적 관계를 검증함.

01.서론

시장 성장과 소비자 불만: 연구의 출발점



비록 전 세계 전기차 시장이 성장을 유지하고 있지만, 소비자들은 배터리 안전, 충전 편리성, 비용 등 여러 측면에서 불만이 계속 증가하고 있음. 이전 연구에서는 세 가지 핵심 불만 요소를 식별했으며, 이를 바탕으로 본 연구에서는 설문 조사를 통해 불만-위험 인식-행동 의도 간의 구조적 관계를 검증함.

01.서론

연구목적

본 연구는 Reddit 플랫폼에서 전기차 관련 부정적인 리뷰의 초기 문서 발굴을 기반으로 소비자 불만의 세 가지 주요 원인을 식별함: 기술 문제, 사용 편의성 문제 및 신뢰 위험.이를 통해 위험을 인식하고 의도를 변화시키기 위한 구조적 모델을 구축하는 것을 목표로 한다. 본 연구는 이러한 불만 요인들이 전기차에 대한 심각성과 감수성에 대한 소비자들의 인식에 어떻게 영향을 미치는지, 그리고 이러한 위험 인식이 전통적인 자동차나 다른 교통수단으로의 전환 의도에 더 영향을 미치는지에 초점을 맞추다. 또한, 본 연구는 "회피 동기"가 위험 인식과 조향 의도 사이에서 조절 역할을 하는지를 실증적으로 분석할 것.

연구질문

전기차의 기술적 측면, 사용성 및 신뢰성에 대한 소비자 불만이 전기차와 관련된 위험(즉, 인식된 심각성 및 인식된 감수성)에 대한 인식에 영향을 미치나요?

인지된 심각성과 인지된 감수성이 소비자들이 전기차를 선택하려는 의도에 큰 영향을 미치나요?

회피하려는 동기가 위험 인식과 변화하는 의도 사이에서 조절 역할을 하나요?

02.이론적 배경

기대 불일치 이론 (Expectation Disconfirmation Theory)

자동차 제조업체가 최대 충전 전력으로 BEV를 충전하는 데 기술적으로 가능한 최단 서비스 시간을 광고하는 경우, 고객은 충전 서비스에 대해 높은 기대치를 갖는 경우가 많음. (L.J. Rosenberg, J.A. Czepiel,1984)

EDT는 고객의 구매 전 기대치와 주관적인 구매 후 평가가 높은 (불)만족도로 이어질 수 있음을 시사. (R.L. Oliver,1980)

건강행동이론(Theories of Health Protective Behaviors)

건강 신념 모델은 처음에는 질병 예방 행동을 설명하기 위해 개발되었지만, 최근에는 위험 회피 행동을 설명하거나 기술을 거부하는 데 사용되고 있다(Godding and Cork, 1996; Champion and Skinner, 2008). 전기차와 관련된 다양한 부정적인 정보(예: 배터리 화재, 주행 중 주차, 충전소 부족 등)는 소비자가 전기차 구매나 사용을 기피하는 결정에 영향을 미칠 수 있다. 이 경우 지각된 심각성과 민감도가 소비자가 접촉을 기피하게 만드는 주요 요인이며, 건강 신념 모델을 적용하면 소비자의 위험 인식과 행동 의도를 구조적으로 설명할 수 있다.

본 연구에서는 EDT와 THPB를 이론적 기초로 하여 '소비자 불만 → 위험 인식 → 회피 동기 → 행동 의도'의 구조적 모델을 구축하고 설문조사를 통해 실증적 검증을 수행.

03.연구가설

Author(s)	Research Content
Liang & Xue (2009); Arachchilage & Love (2014)	TTAT 이론은 사용자가 기술로 인한 위협에 직면할 때, 위협(예: 인지된 심각성 및 인지된 감수성)에 대한 주관적인 평가가 회피 동기를 자극하여 위험을 피하기 위한 조치를 취할 것인지 여부를 결정한다고 제안하다. 이 이론은 정보 보안 및 IT 기술 수용 연구에 널리 적용되어 왔다.
Jain & Agrawal (2020)	소셜 미디어에서 사이버 괴롭힘을 TTAT와 결합하여 연구했다. 온라인에서 발생하는 사이버 범죄이기 때문에 사용되는 IT를 통해 범죄를 피할 수 있는지, 얼마나 심각한 것으로 보이는지, 두려움으로 인한 회피 의지를 확인할 수 있다.
Couraud, 2014; Rhoa & Yub, 2011; Xue et al., 2015n Hippel, 1986	위협 평가의 경우, 지각된 심각도와 지각된 감수성은 회피 동기에 영향을 미칩니다. TTAT는 다양한 IT 연구에 적용되어 왔습니다(Couraud, 2014; Rhoa & Yub, 2011; Xue et al., 2015).
Yixi Xue, Xiaoyan Zhang ,Yi Zhang, Exiang Lu(2024)	소비자들은 전기차의 안전성, 제한된 주행 거리에 대해 우려할 수 있음.

TTAT 이론과 위험 인식에 관한 연구는 사용자가 기술적 위협에 직면했을 때 위험에 대한 주관적인 평가(인식 심각성 및 인식 감수성)가 회피 동기와 행동에 영향을 미친다는 것을 보여준다.

H1a: 기술적 불만은 전기차와 관련된 인지된 심각성을 증가시킬 것이다.
H1b: 사용성 불만은 전기차와 관련된 인지된 감수성을 증가시킬 것이다.
H1c: 신뢰성 불만은 전기차와 관련된 인지된 심각성과 감수성을 증가시킬 것이다.

03.연구가설

Author(s)	Research Content
S. Goel, R. Sharma and A. K. Rathore,(2021)	전기차를 지원하는 적절한 전력망 인프라와 충전소가 부족하다는 점이 판매량 저조, 높은 가격, 침체의 주요 원인 중 하나임.
J. A. Sanguesa, V. Torres-Sanz, P. Garrido, F. J. Martinez and J. M. Marquez-Barja, (2021)	오늘날 전기차에 사용되는 배터리의 현재 용량은 또 다른 주요 과제이다. 저장 어려움과 관련된 비용도 막대하다. 이는 더 높은 에너지 밀도를 달성하기 위한 배터리 기술의 개발과 개선을 통해서만 해결할 수 있는 큰 문제.
전자신문, 2023	정부는 국가보조금 소진율 저하와 이에 따른 전기자동차 보급률 하락에 대응하기 위하여 2023년 9월 전기자동차 구매 보조금(국고분)을 당초 최대 680만원에서 780만원으로 인상하였는데, 전문가들은 이러한 전기자동차 보급 둔화가 고금리와 보조금 감소로 전기차 구매 심리가 위축된 데다 충전 불편과 충전료 인상, 화재 이슈로 소비자 관점에서의 경제성과 편의성 등 전기차 시장 경쟁력이 하락했기 때문으로 보고 있음(전자신문, 2023).
Wang Yuehui, Wang Qing(2013)	장기적인 재정 지원은 전기차 산업에 보다 안정적인 개발 환경을 제공하고 위험을 줄일 수 있기 때문에 보조금 기간을 연장하는 것도 중요한 정책 옵션. 이를 통해 더 많은 투자와 참여를 유도하고 기술 혁신과 시장 발전을 가속화할 수 있음.
배재권. (2017).	지각된 개연성, 지각된 심각성, 지각된 유익성, 지각된 장애성에 대한 신념이 크거나 작을수록 건강을 보호하거나 추구하려는 행동 을 적극적으로 수행한다. 지각된 개연성과 지각 된 심각성에 의해 위협감이 형성되어 행동 가 치가 부여되고, 지각된 유익성과 지각된 장애성 에 의해 그 가치가 평가(행동평가)되어 바람직 한 건강행위 이행으로 나타나게 된다.
Niu Jinfeng, Chen Sichin(2020)	정부는 항상 전기차 산업에 큰 중요성을 부여하고 다양한 문제를 개발하고 추진하는 전기차 산업은 다양한 정책을 지원하기 위해 도입되었지만 정책의 구현이 특히 눈에 띄지 않고 열정이 부족하고 실명.

전기차 산업은 기술, 정책 및 인프라 등 여러 방면에서 연구를 진행하며, 감지된 위험과 불편함이 소비자의 수용 의지를 약화시키고 회피 성향을 강화한다는 것을 보여준다.

H2a : 인지된 심각성은 EV 회피 의도를 증가시킬 것이다.
H2b : 인지된 감수성은 EV 회피 의도를 증가시킬 것이다.

03.연구가설

Author(s)	Research Content
Sheth&Parvatiyar,1995	소비자들이 불만족한 상황에서 어떻게 생각하고 행동하는지를 아는 것은 기업으로 하여금 구매 후 행동 관리,더 나아가서는 소비자와의 관계 형성 및 유지에 큰 도움을 줄 수 있음.
Day(1980)	불만족을 소비의 경험으로 인해 생기는 결과로,행복하지 목한 수준이라고 표현하기도 하고,주관적으로 평가된 기대,제품을 구매하고 소비하는 것 들과 관계되는 여러 가지 경험들을 비교하여 얻는 결과로 해석.
Smith&Bolton,2002	불만족은 고객이 제품이나 서비스를 구매하고 비교 선택하는 과정에서 발생되는 부정적 감정의 경험으로 고객이 만족하지 못한 경우 또는 서비스 제공자가 제공한 상품이나 서비스가 기대에 부응하지 못하다고 느끼는 고객의 감정 상태라고 할 수 있음.
Cecere, G., Corrocher, N. and Guerzoni, M. (2018)	높은 조달 비용이 전기차 보급에 걸림돌이 되는 반면, 낮은 에너지 비용은 전기차 사용의 장점이라고 주장.
Li, L.,Wang, Z., Chen, L., et al. (2020)	제한된 1회 충전 주행거리, 충전소 부족, 배터리 보증 부족과 같은 기술적 문제를 주요 장애물 중 하나로 꼽았음.

기존 연구들은 전기차 사용 과정에서 발생하는 실질적인 불만족이 소비자의 평가에 부정적 영향을 준다는 점을 시사한다.

H3a: 회피 동기는 인지된 위험(심각성 및 감수성)에 의해 증가할 것이다.
H3b: 회피 동기는 EV 회피 의도를 증가시킬 것이다.
H3c: 인지된 위험(심각성 및 감수성)은 회피 동기를 매개로 EV 회피 의도에 영향을 미칠 것이다.
H3d: 인지된 위험하고 회피 동기를 통해 ICEV의 구매 의도와 HEV 구매 의도에 간접적으로 영향을 미칠 것이다.

03.연구가설

Author(s)	Research Content
Wang S., Fan, J., Zhao, D., Yang, S., & Fu, Y. (2016).	환경 인식와 해결 방안, 그리고 환경 인식을 해결하는 방법과 아이디어를 포괄적으로 생각하는 마음가짐이다.
Wang S., Fan, J., Zhao, D., Yang, S., & Fu, Y. (2016).	전기차 도입에 관한 연구에 따르면 환경 문제는 고객의 개인 도덕적 기준, 행동 규범, 개인 행동 통제 및 주관적 규범과 관련이 있는 것으로 나타났다.
Chawla, U.; Mohnot, R.; Mishra, V.; Singh, H.V.; Singh, A.K.2023	연구들은 소비자 환경 인식이 전기차에 대한 선호도와 태도에 미치는 영향을 강조하고 있다.
Chawla, U.; Mohnot, R.; Mishra, V.; Singh, H.V.; Singh, A.K.2023	환경에 대한 인식이 높은 소비자는 TFV에 비해 전기차를 더 친환경적이고 저탄소이며 환경 친화적이라고 인식할 가능성이 높다. 이러한 인식은 기능적 및 심리적 장벽이 구매 의도에 미치는 부정적인 영향을 약화시킨다.

소비자의 환경 인식은 전기차에 대한 태도와 구매 의도에 영향을 미친다.

H4a: 환경 의식은 인지된 심각성이 EV 회피 의도에 미치는 영향을 약화시킬 것이다.
H4b: 환경 의식은 인지된 감수성이 EV 회피 의도에 미치는 영향을 약화시킬 것이다.
H4c: 환경 의식은 위험 인식이 회피 동기 및 EV 회피 의도에 미치는 영향을 약화시킬 것이다.

03.연구가설

Author(s)	Research Content
Li et al. (2015)	14개 국제 도시/지역에서 전기차 도입에 영향을 미치는 요인을 조사한 결과 보조금과 세금 인센티브가 전기차 개발에 필수적이라는 사실을 발견했다.
Wang et al., (2017)	인센티브 정책이 가격을 낮추고 전기차 도입 비용을 절감하여 소비자의 전기차 도입 의도를 촉발할 수 있다는 것은 분명.
Wang et al. (2017)	우대 정책이 전기차 구매 의도에 긍정적인 영향을 미친다고 언급.
Ma et al. (2017)	Ma et al. (2017)은 인센티브 정책의 효과를 정량화하기 위해 다변량 공적분 모델을 설정했으며, 전기차 판매와 인센티브 정책 사이에 긍정적인 공적분 관계가 있음을 발견했다.
Hoen and Koetse (2014)	Hoen과 Koetse(2014)는 인센티브 정책이 네덜란드 소비자들 사이에서 전기차를 도입하려는 의도에는 유용하지만, 전기차의 신뢰성에 대한 불안감을 줄이는 데는 훨씬 덜 효과적이라고 제안했다.

선행 연구에 따르면, 보조금과 세금 혜택 등의 정책 인센티브는 EV 채택 의도(Lietal. , 2015; Wang et al. , 2017; Ma et al. , 2017), 그러나 EV 위험에 대한 소비자의 우려를 완화하는 데는 한계가 있다(Hoen & Koetse, 2014). 따라서 이 연구는 H5a-H5c를 제안하다.

H5a: 정책 인센티브는 인지된 위험(심각성 및 감수성)이 EV 회피 의도에 미치는 영향을 약화시킬 것이다.
H5b: 정책 인센티브는 EV 회피 의도를 약화시킬 것이다.
H5c: 정책 인센티브는 위험 인식이 회피 동기 및 EV 회피 의도에 미치는 영향을 약화시킬 수 것이다.

03.연구가설

Author(s)	Research Content
Buhmann, A., Decker, T., & Mühlbacher, H. (2022)	전기차 가격이 높을수록 전기차에 대한 선호도가 낮아진다.
Buhmann et al., 2022	전기차 가격이 높거나 성능·충전 인프라가 부족할 경우, 소비자는 전기차 대신 기존 내연기관차를 선택한다.
Xing, Leard, & Li, 2019	전기차 보급 연구에서, 상당수 소비자는 여전히 내연기관차 사용을 지속하고 있다.
Aravena et al., 2021	성 기반 선택실험 결과, EV 특성이 매력적이지 않을 때 응답자는 내연기관차를 유지하려는 경향을 보였다.
Hackbarth, A., & Madlener, R. (2013)	소비자는 전기차보다 내연기관차의 낮은 위험성과 친숙성을 더 중시하여 내연기관차를 계속 선택하는 경향이 있다.
Rezvani, Z., Jansson, J., & Bodin, J. (2015)	전기차 채택 장벽 연구에서, 충전 불편과 성능 불확실성은 소비자가 기존 내연기관차 사용을 유지하는 핵심 요인으로 작용한다.

소비자의 가격, 성능 및 충전 등에 대한 불만은 전기차에 대한 선호를 낮추고, 내연기관 차량을 선택하는 경향을 강화할 것이다.

H6: 인지된 위험(심각성 및 감수성)은 내연기관차 구매 의도를 증가시킬 것이다.

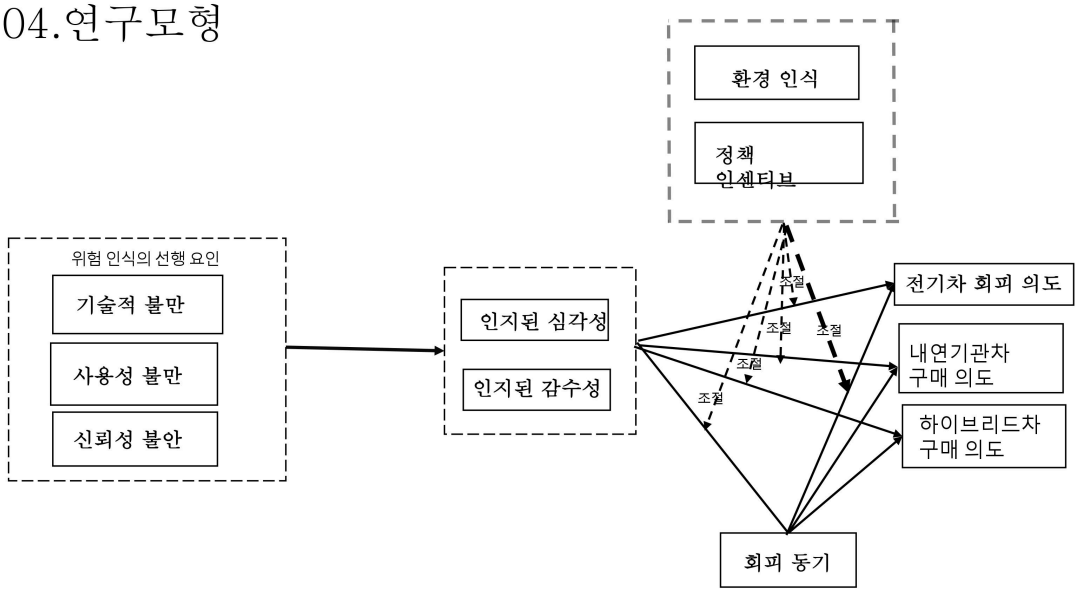
03.연구가설

Author(s)	Research Content
Zamil et al., 2023	하이브리드차는 전기차의 단점을 보완하는 대체 옵션으로, 전기차 위험이 높게 인식될수록 하이브리드차 선택 가능성이 커진다.
Tanwir & Hamzah, 2020	하이브리드차의 구매 의도 연구에서, 환경 태도·사회적 영향이 하이브리드차에 긍정적인 영향을 미쳤다.
Caperello & Kurani, 2012	경적 혜택 인식이 있는 소비자는 ICE 대신 HEV를 우선적으로 고려할 가능성이 높다.
Egbue & Long, 2012	전기차의 충전 불편·주행거리 제약이 커질수록, 하이브리드차 선택 의도가 높아진다 .
Ozaki, R., & Sevastyanova, K. (2011)	하이브리드 차량은 환경적 이미지와 실용적 장점 덕분에 소비자에게 높은 수용 의도를 보여주었다.
Ram, S., & Sheth, J. N. (1989	소비자는 완전한 혁신(전기차)보다 점진적 혁신(하이브리드차)을 더 쉽게 수용하는 경향이 있다.

소비자들은 전기차의 위험이 높다고 느낄 때 하이브리드 차량을 안전과 실용성을 모두 고려한 대체 선택으로 더 선호하다.

H7: 인지된 위험(심각성 및 감수성)은 하이브리드차 구매 의도를 증가시킬 것이다.

04.연구모형



본 연구 모델은 전기차와 관련된 불만 요소(기술, 사용성, 신뢰성)가 소비자의 위험 인식(심각성 및 감수성 인식)을 향상시킬 것이라고 믿다. 위험 인식은 전기차 회피 의도에 직접적인 영향을 미칠 뿐만 아니라, '회피 동기'를 통해 간접적으로 전기차 회피 의도에 영향을 미친다. 또한, 위험 인식은 내연기관 자동차의 구매 의도와 하이브리드 자동차의 구매 의도를 더욱 촉진할 것이다. 동시에 환경 의식과 정책 인센티브가 조절 변수로 작용하여 위험 인식이 전기차 회피 의도에 미치는 영향 경로를 약화시킨다.

05.연구방법

✓ 설문조사

✓ 연구대상(355명): 본 연구의 조사대상은 전기차(EV)에 대한 사용 경험 또는 관련 인식을 보유한 중국 소비자로, 온라인 플랫폼(웨이보·전기차 커뮤니티 등)을 통해 자발적으로 참여한 응답자 355명을 최종 분석에 활용하였다.

특성		빈도	%
성별	남	144	40.6
	여	211	59.4
연령대	18-25세	18	5.1
	26-35세	55	15.5
	36-45세	115	32.4
	46-55세	126	35.5
	56세 이상	41	11.5
월 소득	100만원 미만	28	7.9
	100만원-200원	54	15.2
	200만원-300만원	93	26.2
	300만원-400만원	77	21.7
	400만원-500만원	57	16.1
	500만원-700만원	31	8.7
	700만원이상	15	4.2
EV 보유 여부	예	252	71
	아니요	103	29
운전 경력	1년 미만	135	38
	1-3년	133	37.5
	3-5년	30	8.5
	3-10년	34	9.6
	10년이상	23	6.5

06.결과

Construct (α)		Factor loading	AVE	CR
사용성 불만(0.898)	1	0.815	0.766	0.929
	2	0.819		
	3	0.810		
	4	0.811		
기술적 불만(0.94)	1	0.832	0.77	0.953
	2	0.835		
	3	0.799		
	4	0.841		
	5	0.844		
	6	0.830		
신뢰성 불만(0.925)	1	0.832	0.769	0.943
	2	0.838		
	3	0.799		
	4	0.830		
	5	0.852		
환경 의식(0.883)	1	0.634	0.683	0.915
	2	0.812		
	3	0.842		
	4	0.817		
	5	0.871		
정책 인센티브(0.861)	1	0.652	0.634	0.895
	2	0.774		
	3	0.831		
	4	0.818		
	5	0.826		
인지된 심각성(0.903)	1	0.829	0.771	0.931
	2	0.778		
	3	0.831		
	4	0.803		
인지된 감수성(0.904)	1	0.729	0.777	0.933
	2	0.757		
	3	0.729		
	4	0.770		
회피 동기(0.885)	1	0.834	0.735	0.917
	2	0.788		
	3	0.860		
	4	0.784		
전기차 회피의도(0.846)	1	0.776	0.765	0.907
	2	0.786		
	3	0.836		
	4	0.855		
ICEV 구매 의도(0.895)	1	0.767	0.757	0.926
	2	0.762		
	3	0.767		
HEV 구매 의도(0.865)	1	0.858	0.785	0.916
	2	0.802		
	3	0.805		

05.결과

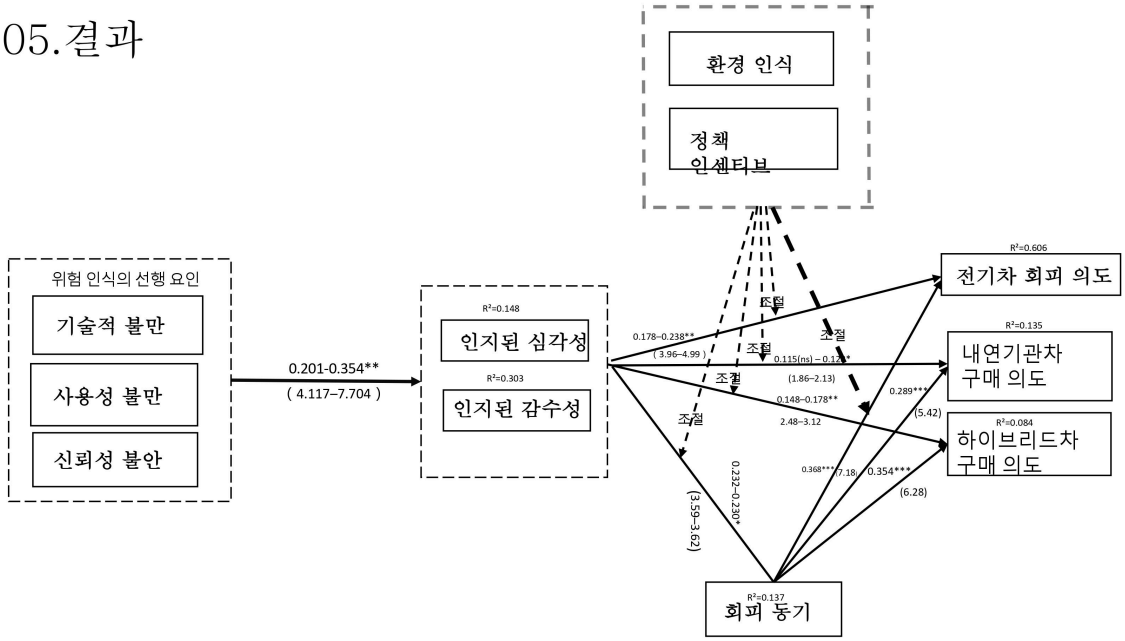
	회피동기	전기차 회피의도	환경 의식	HEV 구매 의도	ICEV 구매 의도	인지된 심각성	인지된 감수성	정책 인센티브	신뢰성 불만	기술적 불만	사용성 불만
회피 동기	**0.857**										
전기차 회피의도	0.336	**0.874**									
환경 의식	-0.191	-0.216	**0.826**								
HEV 구매 의도	0.289	0.301	-0.187	**0.886**							
ICEV 구매 의도	0.368	0.324	-0.196	0.382	**0.87**						
인지된 심각성	0.297	0.355	-0.096	0.306	0.401	**0.878**					
인지된 감수성	0.296	0.409	-0.432	0.381	0.428	0.284	**0.881**				
정책 인센티브	-0.149	-0.142	0.249	-0.062	-0.162	-0.367	-0.083	**0.796**			
신뢰성 불만	0.299	0.316	-0.171	0.347	0.373	0.336	0.432	-0.043	**0.877**		
기술적 불만	0.392	0.349	-0.176	0.354	0.46	0.294	0.408	-0.089	0.349	**0.877**	
사용성 불만	0.305	0.392	-0.179	0.371	0.397	0.361	0.442	-0.072	0.262	0.324	**0.875**

✓ 상관 행렬 분석 결과, 각 변수의 상관 계수가 대각선에 있는 해당 변수의 AVE의 산술 제곱근보다 작아 변수를 잘 구분할 수 있으며, 서로 다른 변수들이 혼동되지 않음을 보여주었으며, 이는 본 연구의 설문지가 좋은 판별 타당성을 가지고 있음을 증명한다.

05.결과

가설	원본 샘플(O)	T통계(|O/STDEV)	P 값
기술적 불만 → 인지된 심각성(H1a) ✓	0.201	4.117	0.000
사용성 불만 → 인지된 감수성(H1b) ✓	0.354	7.704	0.000
신뢰성 불만 → 인지된 심각성/감수성 (H1c) ✓	0.266–0.339	5.037–6.954	0.000
인지된 심각성 → EV 회피 의도 (H2a) ✓	0.178	3.962	0.000
인지된 감수성 → EV 회피 의도(H2b) ✓	0.148	3.179	0.001
인지된 위험 → 회피 동기(H3a) ✓	0.232–0.230	4.565–4.542	0.000
회피 동기 → EV 회피 의도(H3b) ✓	0.354	6.280	0.000
인지된 위험 → 회피 동기 매개 → EV 회피 의도(H3c) ✓	0.027–0.050	2.367–2.902	0.004-0.018
인지된 위험 → 회피 동기 매개 → ICEV/HEV 구매 의도(H3d) ✓	0.067–0.085	3.395–3.729	0.000-0.001
환경 의식 × 인지된 심각성 → EV 회피 의도(H4a) ✓	-0.109	1.987	0.047
환경 의식 × 인지된 감수성 → EV 회피 의도(H4b) ✓	-0.146	2.316	0.021
환경 의식 × 인지된 위험 → 회피 동기 및 EV 회피 의도(H4c) ✓	-0.128	2.102	0.036
정책 인센티브 × 인지된 위험 → EV 회피 의도(H5a) ✓	-0.115	2.214	0.027
정책 인센티브 × 회피 동기 → EV 회피 의도(H5b) ✓	-0.109	1.987	0.047
정책 인센티브 × 인지된 위험 → 회피 동기 및 EV 회피 의도(H5c) ✓	-0.118	2.145	0.033
인지된 위험 → ICEV 구매 의도(H6) ✓	0.085	3.729	0.000
인지된 위험 → HEV 구매 의도(H7) ✓	0.067	3.264	0.001

05.결과



07.결론

- ✓ 본 연구는 전기차 소비자의 불만 요인(기술적·사용성·신뢰성 불만) 이 인지된 위험(심각성 및 감수성), 회피 동기, 그리고 구매 관련 행동(회피 의도, ICEV 및 HEV 구매 의도) 에 미치는 영향을 실증적으로 분석하였다.연구 결과, 대부분의 가설이 지지되었으며 주요 변수 간에 유의미한 연관성이 있음을 확인하였다.
- ✓ 분석 결과, 기술적·사용성·신뢰성 불만은 모두 인지된 심각성과 감수성을 유의하게 증가시켰으며,인지된 위험은 회피 동기를 강화하여 EV 회피 의도로 이어지는 것으로 나타났다.이는 소비자 불만이 단순한 감정적 반응을 넘어 행동적 결과로 연결된다는 점을 시사한다.또한 회피 동기는 내연기관차(ICEV) 및 하이브리드차(HEV) 구매 의도에도 유의한 매개 효과를 보였다.
- ✓ 환경 의식과 정책 인센티브는 조절 변수로서 위험 인식의 부정적 영향을 완화시키는 역할을 하였으며,정책 인센티브는 EV 회피 의도를 약화시키는 유의한 효과를 보였다.
- ✓ 본 연구는 소비자 불만을 '시장 이탈의 원인'이 아닌 '혁신의 단서'로 해석할 수 있음을 보여준다.이에 따라 전기차 제조사와 정책 입안자는 부정적 리뷰를 제품·서비스 혁신을 위한 피드백 자원으로 활용할 필요가 있으며,이를 통해 전기차 시장의 신뢰 회복과 지속 가능한 발전을 도모할 수 있을 것이다.

07.결론

학술적 시사점

- ✓ 본 연구의 학문적 의의는 소비자 불만을 전기차 시장의 위험 인식 및 회피 행동과 연결하여 검증했다는 점이다.
- ✓ 기존 연구들이 전기차 수용 의도나 기술 혁신에 초점을 두었던 것과 달리, 본 연구는 불만(기술적·사용성·신뢰성 요인) 을 인지된 심각성과 감수성으로 구조화하여 소비자 위험 인식 과정의 구체적 경로를 실증적으로 제시하였다.이는 불만이 단순한 부정적 감정이 아니라 위험 지각을 자극하는 심리적 기제임을 보여주었으며,소비자 회피 행동 연구에서 인지된 위험과 회피 동기의 매개적 역할을 명확히 규명했다는 점에서 학문적 가치가 있다.
- ✓ 환경 의식과 정책 인센티브의 조절 효과를 실증적으로 확인함으로써, 외적 요인이 소비자 행동의 완충 장치로 작용함을 제시하였다.따라서 본 연구는 불만-위험 인식-행동 간의 연계 구조를 통합적으로 검증한 확장된 모형을 제시함으로써,소비자 심리 및 혁신 수용 연구에 새로운 시각적 틀을 제공하였다.

실무적 시사점

- ✓ 본 연구의 결과는 전기차 제조사와 정책 입안자에게 실질적인 전략적 인사이트를 제공한다.
- ✓ 첫째, 소비자 불만은 시장 이탈의 신호가 아니라 혁신의 출발점으로 인식되어야 한다.기업은 소비자의 기술적·사용성·신뢰성 불만 데이터를 분석하여 제품 성능 개선, 충전 인프라 확충, 안전성 강화 등 서비스 혁신의 방향을 설정할 수 있다.
- ✓ 둘째, 인지된 위험이 회피 행동으로 이어지는 과정에서 환경 의식과 정책 인센티브가 완충 역할을 하는 만큼,정부와 산업계는 소비자의 신뢰 회복을 위한 인센티브 설계 및 친환경 가치 홍보 전략을 강화할 필요가 있다.
- ✓ 마지막으로, 온라인 부정 리뷰를 단순히 통제 대상이 아닌 '소비자 목소리 기반의 혁신 자원'으로 전환할 때,전기차 시장의 지속 가능한 성장과 브랜드 신뢰 제고가 가능할 것이다.

	Memo	

Exploring the Expectation-Mismatch in Electric Vehicle Adoption: Insights from Negative Consumer Reviews on Reddit

정자강
순천향대학교 경영학과
–
왕문준
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

As the global electric vehicle (EV) market transitions from rapid growth to stable development, consumer complaints are increasingly evident in negative social media comments. These comments serve not only as critical post-adoption feedback but also as drivers of technological and service innovation. This study analyzes negative EV reviews to identify key factors underlying consumer dissatisfaction and to explore implications for market acceptance and innovation strategies. Integrating expectation–mismatch theory with social media marketing theory, the study interprets these findings theoretically while demonstrating the value of text mining for capturing post-adoption consumer experiences. Practical recommendations are provided for product optimization, service improvement, and public opinion management in the EV sector.

2025 한국지식경영학회 추계학술대회

타이어 O2O 서비스에서 고객 만족과 충성도의 교차효과 분석

An Analysis of Cross-over Effects Between Customer Satisfaction and Loyalty in Tire O2O Services

2025. 11. 13

김상준(경희대학교 일반대학원 경영학과 박사과정, sj.kim@khu.ac.kr)
양성병(경희대학교 일반대학원 경영학과 & 빅데이터응용학과 교수, sbyang@khu.ac.kr)

KYUNG HEE UNIVERSITY

목 차

Ⅰ. 서론
Ⅱ. 이론적 배경
Ⅲ. 연구 모형 및 가설
Ⅳ. 예상 결론 및 시사점

부록: 설문지의 구성 / 참고문헌

1

Ⅰ. 서론

1-1. 연구의 배경

- 타이어는 차량 **안전과 직결**되는 **제품과 장착서비스가 결합된 '고관여'** 품목으로, 소비자들은 구매 시 다양한 정보를 탐색하고, 민감하게 반응한다(현명진 외, 2020).
- 플랫폼 기반 상거래가 성숙 단계에 접어들면서, 한국 시장에서도 **타이어의 온라인 유통 채널의 비중이 꾸준히 커지고 있다** (IMARC Group, 2024).
- 타이어 유통 시장은 고객들이 온라인 플랫폼에서 정보를 탐색하고 구매한 뒤, 실제 교체는 오프라인 매장에서 수행하는 **전형적인 O2O 구조를 따른다**(최성국 외, 2020).
- 최근 연구들은 이러한 **O2O 구매 여정(O2O customer journey)이 고객 만족과 충성 형성에 새로운 메커니즘을 제시**한다고 보고 있다(Verhoef et al., 2015).
- 특히 온라인 단계에서의 구매 만족이 오프라인 서비스 만족 및 재이용 의도에 영향을 미치는 등, **채널 간 만족과 충성도의 교차효과(cross-over effect)**가 실증적으로 확인되고 있다(장이월 외, 2022; Ayachi et al., 2024).

Ⅰ. 서론

1-2. 연구의 필요성

- O2O(Online to Offline) 서비스는 온라인에서의 정보 탐색 및 구매와 오프라인에서의 체험 및 소비가 결합된 형태로, **고객의 경험이 한 채널에 국한되지 않고 상호작용하는 특징을 가진다**(Verhoef et al., 2015).
- 특히 고관여 제품의 경우, 온라인 구매 경험이 오프라인 서비스 만족으로, 혹은 오프라인 서비스 경험이 온라인 플랫폼 충성으로 확장되는 **교차효과(cross-over effect)**가 발생할 수 있다(장이월 외, 2022).
- 최근 연구들에 따르면, 이러한 교차효과는 O2O 환경에서 고객충성도를 설명하는 핵심 요인으로 작용하며(Ayachi et al., 2023), 온라인과 오프라인 채널 간의 통합 수준이 높을수록 고객 만족과 재이용 의도가 강화되는 것으로 나타났다.
- 그러나 지금까지의 연구는 대부분 숙박, 음식배달, 유통 등 저관여 서비스 중심의 O2O 맥락에 국한되어 있다(Verhoef et al., 2015; Zhang et al, 2022). 반면, 타이어처럼 제품 품질(안전성과 기능성)과 장착서비스(기술력, 신뢰성)가 동시에 평가되는 **고관여 O2O 제품군에서는 이러한 교차효과가 어떻게 작용하는지에 대한 실증적 연구가 부족**하다.
- 따라서 본 연구는 타이어 O2O 서비스라는 고관여 제품군을 대상으로, 온라인 및 오프라인 만족이 서로의 충성도로 확장되는 **교차효과를 규명함으로써** 기존 O2O 연구의 한계를 보완하고, 플랫폼 기반 자동차 서비스 산업의 고객경험 관리 전략에 학문적·실무적 시사점을 제시할 필요가 있다.

Ⅰ. 서론

1-3. Research Gap

- 첫째, **기존 O2O 연구의 산업적 범위 한계**가 존재한다.
  기존 연구는 대부분 **저관여 서비스 산업(숙박, 외식, 유통 등)** 에 초점을 맞추고 있으며(Verhoef et al., 2015; Zhang et al., 2022), 타이어처럼 제품과 서비스가 결합된 **고관여 제품군(O2O 결합형 제품)** 에 대한 실증적 분석은 거의 이루어지지 않았다.

- 둘째, **기존 연구의 단일 채널 중심 분석 한계**가 지적된다.
  O2O 서비스의 본질은 온라인과 오프라인의 결합임에도 불구하고, 대다수 선행연구는 '온라인 만족→온라인 충성' 혹은 '오프라인 만족→오프라인 충성'과 같은 단일 경로 모델에 머물러 있다(최성국 외, 2020).

- 셋째, **타이어 O2O 플랫폼의 맥락적 특수성이 기존 연구에서 간과**되었다.
  타이어는 차량 안전성과 직결되는 고위험 제품으로, 구매 의사결정이 합리적 판단과 정보 탐색에 기반한다(현명진 외, 2020). 이러한 특성은 일반적 서비스와 달리, 고객이 온라인 플랫폼의 정보 품질뿐 아니라 오프라인 매장의 기술 신뢰도와 서비스 품질을 동시에 고려하도록 만든다. 따라서 타이어 O2O 시장은 교차효과를 검증하기에 적합한 산업적 맥락임에도, 해당 영역을 다룬 연구는 아직 발견되지 않는다.

Ⅰ. 서론

1-4. 연구 목적

이에 본 연구는 타이어 O2O 플랫폼 이용 고객을 대상으로,

- **온라인 구매 만족이 오프라인 장착서비스 만족 및 충성도에 미치는 영향,**
- **오프라인 서비스 만족이 온라인 플랫폼 충성도에 미치는 영향,**
- **만족도 간 상호관계에 의해 형성되는 교차효과(cross-over effect)** 를 실증적으로 검증하고자 한다.

이를 통해, 본 연구는 기존 O2O 연구가 간과해 온

① 고관여 제품의 복합적 경험 구조,

② 온라인–오프라인 간 만족 및 충성도의 교차관계,

③ 선택속성 기반 고객경험 메커니즘을 통합적으로 분석함으로써,

플랫폼 기반 자동차 서비스 산업의 고객충성도 형성과 관리 전략에 학문적·실무적 시사점을 제시하고자 한다.

Ⅱ. 이론적 배경 및 선행연구

2-1. 타이어 산업 동향

- **2024년 타이어 산업의 전체 판매량은 8,801만개로 전년 대비 4.7% 증가하였다.** 이러한 성장세는 내수 판매와 수출에서 서로 다른 양상을 보이며 전개될 것으로 예상된다. 내수 판매를 살펴보면, 국내 완성차 산업의 생산 및 수요 감소로 신차용(OE: Original Equipment) 판매가 크게 둔화(-13.2%)될 것으로 보인다. 그러나 교체용(RE: Replacement Equipment) 판매가 견조한 성장(+4.7%)을 나타내고 있다(대한타이어협회, 2025)

승용차용 타이어 판매 현황(단위: 본)

구분	2022	2023	2024	2025(1~5월)
신차용	4,597	3,857	3,520	1,497
교체용	**12,527**	**13,044**	**13,670**	**5,222**

※출처: 대한타이어산업협회(2025)

Ⅱ. 이론적 배경 및 선행연구

2-2. 플랫폼 기반 타이어(자동차 및 자동차용품) 구매 동향

- 최근 플랫폼 기반 상거래에서 자동차 및 자동차용품 상품군의 판매는 2024년 총 거래액 5조 8,046억 원으로 전년 4조 8,065억 원에 비해 **20.76% 증가하였다**(국가데이터처, 2025).
- 이 중에서 국내 타이어 시장규모는 내수 판매가 약 1조 8,000억 원이며, 국내 타이어 시장의 점유율을 보면 한국타이어가 47%, 금호타이어가 38.5%, 넥센타이어가 9.3%, 수입타이어가 5.1%를 차지하고 있다. 특히 한국타이어와 금호타이어가 전체 시장의 85% 이상을 점유하며 시장을 주도하고 있다(대한타이어산업협회, 2024).

(단위: 억원)

	생 활			서비스						기 타
	자동차 및 자동차용품	가구	애완용품		여행 및 교통서비스	문화 및 레저서비스	이쿠폰 서비스	음식서비스	기타서비스	
2022	42,601	51,950	22,733	609,119	168,165	24,501	75,354	316,369	24,729	34,994
2023	48,065	53,521	25,893	766,811	293,920	30,069	97,423	323,722	21,676	35,497
2024	58,046	56,670	27,807	841,770	324,647	31,617	86,518	369,891	29,096	28,889

※상품군별 온라인 쇼핑거래액(국가데이터처, 2025)

Ⅱ. 이론적 배경 및 선행연구

2-3. 교차 효과(Crossover effect)

- **교차효과(crossover effect)**는 본래 서비스 시스템 간 상호작용에서 발생하는 심리적·행동적 전이 현상으로, 한 서비스 제공 과정에서 형성된 고객의 평가나 감정이 다른 서비스나 접점으로 확장되는 것을 의미한다(Smith et al., 1995).

- 즉, 교차효과는 서로 다른 서비스 주체 간의 경험이 만족도와 충성도 형성에 상호 영향을 미치는 내적 비대칭 관계를 설명하는 개념이다(Chu et al., 1995; Mittal et al., 1999)

- 최근 연구에서는 모바일 플랫폼과 오프라인 플랫폼 간 소비 기반의 진화적 상호작용 속에서 교차효과가 점차 강화되고 있음을 실증하였다. 그들은 교차효과가 특정 서비스 제공자의 개별 경험에 국한되지 않고, 상호 연계된 플랫폼 생태계 전체에서 고객경험이 통합적으로 인식되는 과정에서 나타난다고 설명하였다. 이 연구는 고객 만족과 충성도 형성이 단일 채널에서 독립적으로 작동하지 않고, 서로 다른 채널의 평가가 상호 영향을 주는 O2O 플랫폼 경험 구조로 확장됨을 보여주었다(Ha et al., 2020).

- 이러한 교차효과 개념은 O2O 서비스 환경에서도 유효하다. 온라인과 오프라인 서비스는 서로 다른 물리적·디지털 채널임에도 불구하고, 고객은 이를 단일한 **통합경험(Integrated Experience)** 으로 인식하기 때문이다(Ha et al., 2020).

8

Ⅱ. 이론적 배경 및 선행연구

2-4. O2O(온라인) 선택속성과 온라인 서비스 만족도

연구 맥락	플랫폼 품질 (웹 효율성·정보 품질)	배송 서비스 (배송속도·정확성)	지각된 가격 (공정성·합리성)	예약 서비스 (편리성·신뢰성)	연구자
온라인 여행 예약 플랫폼	✓		✓	✓	Jou et al.(2021)
e-서비스 품질 (Quality Dimensions)	✓		✓		Amin et al.(2013)
온라인 숙박 플랫폼	✓	✓		✓	Ju et al.(2019)
전자상거래(e-commerce)	✓	✓	✓		Onur et al.(2021)
모바일 앱 서비스	✓		✓	✓	전상민(2019)
온라인 리테일 서비스	✓	✓			김상태 외(2017)
O4O 플랫폼(허마셴셩)	✓	✓	✓		최성국 외(2020)
타이어 O2O 플랫폼 온라인 단계	✓	✓	✓	✓	**본 연구**

9

Ⅱ. 이론적 배경 및 선행연구

2-5. O2O(오프라인) 선택속성과 오프라인 매장 만족도

연구 맥락	접근성 (접근 용이성·위치 편의성)	매장시설 (물리적 환경·편의성)	직원서비스 (전문성·친절도)	연구자
호텔 숙박시설 만족 요인	✓	✓	✓	Sohrabi et al.(2012)
호텔 서비스 품질	✓	✓	✓	Qu et al.(2000)
쇼핑몰 서비스 품질 및 고객만족	✓	✓	✓	El-Adly(2019)
호텔산업 오프라인 서비스 품질	✓	✓	✓	전상민 (2019)
O4O 매장서비스(허마셴셩)		✓	✓	최성국 외(2020)
자동차 서비스센터 고객만족	✓	✓	✓	Aydin et al.(2023)
타이어 O2O 오프라인 장착서비스	✓	✓	✓	**본 연구**

- **만족도**는 고객이 제품이나 서비스를 이용하면서 사전 기대(expectation)와 실제 성과(performance)를 비교한 결과로서, 그 차이에 대한 인지적·감정적 평가를 의미한다(Oliver, 1997).
- 즉, 제공된 서비스가 기대 이상으로 충족될 때 긍정적 감정이 형성되고, 이는 **충성도(재이용, 추천, 긍정적 구전)의 선행 요인으로 작용한다**(Fornell, 1992; Kim et al., 2004).

10

Ⅱ. 이론적 배경 및 선행연구

2-6. 선행연구 고찰(1/2) – 단일 경로 연구

연구주제	관여도	채널	연구 내용 요약	저자
O2O 통합경험과 만족·충성도 관계	저관여(리테일)	멀티채널(교차채널)	온라인과 오프라인의 채널 통합이 고객만족과 재이용 의도에 긍정적 영향을 미침.	Verhoef et al.(2015)
호텔 선택속성 → 고객만족	저관여(호텔)	단일채널(오프라인)	접근성, 시설, 직원서비스가 호텔 고객만족에 직접적 영향.	Sohrabi et al.(2012)
온라인 예약서비스 품질 → 만족도	저관여(숙박예약)	단일채널(온라인)	예약 절차의 편리성, 신뢰성 등이 온라인 만족에 영향.	Ju et al.(2019)
e-서비스 품질 → 만족도	저관여(전자상거래)	단일채널(온라인)	웹사이트 정보성·신뢰성·반응성이 만족의 핵심 요인으로 검증됨.	Zeithaml et al.(2002)
호텔예약사이트 선택속성 → 만족·재이용	저관여(호텔예약)	단일채널(온라인)	사용 편리성, 가격, 정보 신뢰성이 만족 및 재예약 의도에 영향.	전상민(2019)
O4O 플랫폼 속성 → 만족·충성도	저관여(신선식품)	멀티채널(O4O)	모바일앱 품질, 결제, 상품품질, 매장시설이 만족에 영향, 만족은 충성도로 이어짐.	최성국 외(2020)
온라인 서비스 품질 → 만족도	저관여(의류·소비재)	단일채널(온라인)	웹사이트 디자인·정보 정확성·반응성이 만족에 긍정적 영향.	Kim et al.(2004)
타이어 구매의도 결정요인	고관여(타이어)	단일채널(오프라인)	관여도, 가격, 브랜드, 안전 인식이 타이어 구매의도에 영향.	현명진 외(2020)
자동차 서비스센터 만족 요인	고관여(자동차 A/S)	단일채널(오프라인)	접근성, 매장시설, 직원서비스가 만족과 재방문 의도에 유의한 영향.	Aydin et al.(2023)
쇼핑몰 물리적 환경 → 만족·충성	중관여(쇼핑몰)	단일채널(오프라인)	매장시설, 접근성, 분위기 등 물리적 환경이 만족과 충성도에 영향.	El-Adly et al.(2019)
타이어 교체행동(정보비대칭)	고관여(타이어)	단일채널(오프라인)	소비자 정보 부족과 전문가 의존으로 구매 의사결정 복잡성이 큼.	Mahama-Musah et al.(2020)

11

Ⅱ. 이론적 배경 및 선행연구

2-6. 선행연구 고찰(2/2) - 교차 경로 연구

연구주제	관여도	채널	연구 내용 요약	저자
O2O 숙박서비스 만족 ↔ 충성 교차효과	저관여(숙박공유)	멀티채널(O2O)	온라인 만족이 오프라인 충성으로, 오프라인 만족이 온라인 충성으로 확장됨.	Zhang et al.(2022)
채널 통합 → 만족·충성도	저관여(유통)	멀티채널(Omni)	교차채널 통합 수준이 높을수록 만족과 충성도가 강화됨.	Ayachi et al.(2024)
온라인↔오프라인 신뢰·만족 전이	저관여(일반 서비스)	멀티채널(O2O)	온라인 만족이 오프라인 신뢰 및 충성도로 전이되는 교차효과 실증.	Ha et al.(2020)

- 기존 연구는 대부분 **저관여 서비스(숙박·리테일 등)를 대상으로 한 단일채널 중심의 만족도–충성도에 관한 연구**였다.

- 반면, 본 연구는 고관여 제품인 타이어 O2O 산업을 대상으로 **온라인과 오프라인 간의 양방향 교차효과(cross-over effect)를 검증함**으로써 고관여 제품·서비스 결합형 O2O 고객경험(고객 만족과 충성의 관계)을 보다 종합적으로 이해함으로써 이론적/실무적 시사점을 제공하고자 한다.

12

Ⅲ. 연구방법

3-1. 연구 방법

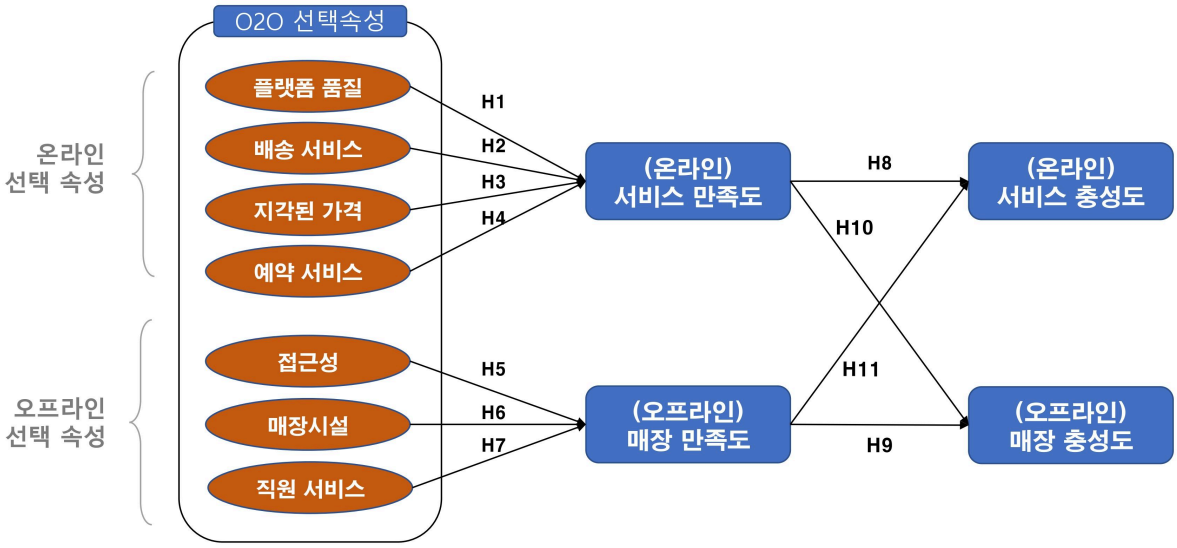
- 본 연구는 국내(전국)에서 **최근 6개월 이내 플랫폼 상거래에서 타이어 구매 경험자를 대상으로 한다. 설문은 온라인 설문지를 활용하여 실시할 예정**이며, 목표 표본수는 약 300부이다. 사전조사(Pilot)로 30명을 실시한 후 문항을 보완하여 본 설문을 실시한다.

- 이는 구조방정식 적용 시 적절한 표본수 확보를 위한 것으로, 일반적으로 자유도 대비 10배 이상 혹은 200~300명 수준이 권장되고 있는 표본의 수로써 구조방적식 모형을 이용하여 진행할 예정이다.
  - ✓ 첫째 각 변수별 문항에 대해 Cronbach's α를 산출해 내부신뢰도를 확인한다.
  - ✓ 둘째 **확인적 요인분석(CFA) 및 측정 모델분석을 통한 연구모델 적합도를 평가하고**,
  - ✓ 셋째 각 변수의 수렴타당도 및 판별타당도를 점검하며,
  - ✓ 넷째 **연구모델 가설 검정 및 매개효과를 분석한다**.
  - ✓ 다섯째 순차적 다집단 분석으로 조절효과를 규명한다.

13

Ⅲ. 연구모형 및 가설

3-1. 연구모형



Ⅲ. 연구모형 및 가설

3-2. 연구가설

Ⅰ. 온라인 선택속성과 온라인 서비스 만족도 관계

H1. 플랫폼 품질은 온라인 서비스 만족도에 정(+)의 영향을 미칠 것이다.
H2. 배송 서비스는 온라인 서비스 만족도에 정(+)의 영향을 미칠 것이다.
H3. 지각된 가격은 온라인 서비스 만족도에 정(+)의 영향을 미칠 것이다.
H4. 예약 서비스는 온라인 서비스 만족도에 정(+)의 영향을 미칠 것이다.

Ⅱ. 오프라인 선택속성과 오프라인 매장 만족도 관계

H5. 접근성은 오프라인 매장 만족도에 정(+)의 영향을 미칠 것이다.
H6. 매장시설은 오프라인 매장 만족도에 정(+)의 영향을 미칠 것이다.
H7. 직원서비스는 오프라인 매장 만족도에 정(+)의 영향을 미칠 것이다.

Ⅲ. 만족도와 충성도 간의 관계

H8. 온라인 서비스 만족도는 온라인 충성도에 정(+)의 영향을 미칠 것이다.
H9. 오프라인 매장 만족도는 오프라인 충성도에 정(+)의 영향을 미칠 것이다.

Ⅳ. 만족도의 교차효과(Cross-over Effect)

H10. 온라인 서비스 만족도는 오프라인 매장 충성도에 정(+)의 영향을 미칠 것이다.
H11. 오프라인 매장 만족도는 온라인 충성도에 정(+)의 영향을 미칠 것이다.

Ⅲ. 연구모형 및 가설

3-3. 변수의 조작적 정의

변수			조작적정의	선행연구
O2O 선택 속성	온라인 선택속성	플랫폼 품질	고객이 타이어 O2O 플랫폼을 이용하는 과정에서 느끼는 웹사이트/앱의 사용 편의성, 응답속도, 정보의 정확성과 신뢰성 수준	Zeithaml et al.(2002); Kim et al.(2004); 최성국 외(2020)
		배송 서비스	온라인 주문 후 제품 배송 과정의 신속성, 정확성, 안전성에 대한 고객의 인식 수준	Brady et al.(2002); Grönroos (1984); 최성국 외(2020)
		지각된 가격	고객이 인식하는 타이어 제품의 가격이 합리적이고 공정하다고 느끼는 정도	Monroe(1973); Westbrook(1981); 최성국 외(2020)
		예약 서비스	온라인 플랫폼에서 예약·결제·스케줄 관리 과정이 얼마나 간편하고 신뢰성 있게 이루어지는지에 대한 인식	Ju et al.(2019); Jou et al.(2021); Verhoef et al.(2015)
	오프라인 선택속성	접 근 성	고객이 오프라인 장착점을 방문하기까지의 물리적 거리, 위치 편리성, 교통 접근성에 대한 인식	El-Adly(2019); Sohrabi et al.(2012); Aydin et al.(2023)
		매장 시설	오프라인 매장의 물리적 환경, 청결도, 장비 수준, 고객 편의시설 등에 대한 인식	Parasuraman et al.(1988); Qu et al.(2000); 최성국 외(2020)
		직원 서비스	직원의 응대 태도, 전문성, 친절성, 문제 해결 능력 등 인적 서비스 품질에 대한 인식	Parasuraman et al.(1988); Aydin et al.(2023); Sohrabi et al.(2012)
만족도		온라인 서비스 만족도	온라인 플랫폼 이용 전반(정보 탐색, 예약, 결제, 배송 등)에 대한 기대 대비 성과의 인지적·정서적 만족 수준	Oliver(1997); Anderson et al.(2003); 최성국 외(2020)
		오프라인 매장 만족도	오프라인 매장에서의 장착서비스, 시설, 직원응대 등에 대한 전반적인 만족 수준	Parasuraman et al.(1988); Fornell (1992); Aydin et al.(2023)
충성도		온라인 서비스 충성도	온라인 플랫폼에 대한 재이용 의도, 타인 추천의도, 긍정적 구전 등 행동적·태도적 몰입 수준	Oliver(1999); Dick et al.(1994); Zhang et al.(2022)
		오프라인 매장 충성도	오프라인 매장(장착점)에 대한 재방문 의도, 서비스 재이용 의도, 추천 의도 수준	Fornell(1992); Aydin et al.(2023); 최성국 외(2020)

16

Q & A

감사합니다.

Track
4-4

브랜드 스토리와 AI 인플루언서

좌장: 윤상혁 동국대학교 교수

스토리의 힘: 디지털 시대 브랜드 스토리텔링이 소비자–브랜드 간 심리적 유대를 형성하는 메커니즘

유몽(순천향대학교)
최재원(순천향대학교)

Exploring User Acceptance of AI Influencers: A Big Data Analysis of Trust, Engagement, and Brand Influence through Instagram

시함(순천향대학교)
최재원(순천향대학교)

Automatic Detection Thumbnail–Title Congruence in YouTube Branded Videos using Multimodal Machine Learning: Implications for View-Based Engagement

르키아(순천향대학교)
최재원(순천향대학교)

스토리의 힘: 디지털 시대 브랜드 스토리텔링이 소비자-브랜드 간 심리적 유대를 형성하는 메커니즘

유몽
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

본 연구는 디지털 시대의 브랜드 스토리텔링 광고를 대상으로, 처리 유창성 과 내러티브 신뢰성이 소비자의 자기 연관성, 내러티브 몰입, 그리고 자기 일치성을 매개로 하여 브랜드 자산과 구전 의도에 미치는 영향을 분석하였다. 또한, 디지털 리터러시와 자기 해석을 개인차 변수로 설정하여, 이들이 서사 광고 효과의 심리적 경로에 미치는 조절적 역할을 검증하였다.

연구 설계는 실험 노출과 설문조사를 결합한 양적 방법으로 구성되었다.참가자들은 처리 유창성(높음/낮음) 과 내러티브 신뢰성(높음/낮음) 의 조합에 따라 네 가지 유형의 유튜브 브랜드 스토리텔링 광고 중 하나를 무작위로 시청하였다. 시청 후, 광고에 대한 인지적·정서적 반응(자기 연관성, 내러티브 몰입, 자기 일치성) 및 브랜드 관련 결과(브랜드 자산·구전 의도)를 측정하였다. 온라인·오프라인을 통해 총 500부의 설문을 수집하였으며, 불완전 응답을 제외한 360부(유효 응답률 72%) 가 최종 분석에 사용되었다.

연구 1에서는 SmartPLS 4.0을 이용하여 구조방정식모형을 검증한 결과, 처리 유창성과 내러티브 신뢰성은 모두 자기 연관성에 유의한 정(+)의 영향을 미쳤으며, 자기 연관성은 내러티브 몰입과 자기 일치성을 매개하여 브랜드 자산 및 구전 의도에 긍정적인 영향을 주는 연속적 매개 경로가 확인되었다. 또한 디지털 리터러시는 처리 유창성과 자기 연관성 간의 관계를 강화하는 조절 효과를 보였다.아울러 연구 1에서는 자기 해석을 전체적 성향으로 포함하여 그 조절효과를 함께 검증하였다.분석 결과, 자기 구해석은 자기 일치성과 브랜드 자산 간의 관계에서 일정 부분 조절효과를 보였다.다만, 이 단계에서는 자기 구성의 하위 유형(독립형/상호의존형)을 구분하지 않고 전체적 특성으로 분석하였다.

연구 2에서는 동일한 데이터셋을 활용하되, 자기 구성을 독립형과 상호의존형으로 세분화하여 보다 정교한 조절효과를 추가적으로 검증하였다. PROCESS Macro(Model 1) 분석 결과, 자기 일치성과 브랜드 자산 간의 관계에서는 유의한 조절효과가 확인되었으나, 구전 의도 경로에서는 통계적으로 유의하지 않은 결과가 나타났다.즉, 자기 해석의 하위 유형은 소비자의 브랜드 평가(인지적·정서적 반응)에는 영향을 미치지만, 외현적 행동(구전 의향)에는 직접적 차이를 보이지 않았다.

본 연구는 내러티브 설득 이론(Narrative Persuasion Theory) 과 자극–유기체–반응(S–O–R) 모델 을 통합하여 디지털 시대 브랜드 커뮤니케이션의 심리적 작동 메커니즘을 제시함으로써, 학문적 확장성과 실무적 시사점을 동시에 제공한다.

KEYWORDS 브랜드 스토리텔링, 처리 유창성, 내러티브 신뢰성, 자기 연관성, 자기 일치성, 디지털 리터러시, 자기 구성, 브랜드 자산, 구전 의도

Exploring User Acceptance of AI Influencers: A Big Data Analysis of Trust, Engagement, and Brand Influence through Instagram

시함
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

This study examines the growing role of AI-powered virtual influencers in social media marketing and their impact on consumer perceptions, and behaviors. Despite the rapid expansion of this phenomenon research exploring how virtual influencers shape trust, engagement loyalty and brand relationships remains limited, using a big data approach this study analyzes more than 429,193 Instagram comments through sentiment analysis multiple regression, and structural equation modeling (SEM) to explore how key influencer attributes such as credibility attractiveness, and content quality affect consumer responses. The research aims to provide a comprehensive understanding of how users interact with AI-generated influencers and how these interactions influence digital relationship building. Overall, the study offers valuable insights for scholars and marketers seeking to better understand and effectively implement AI influencers contemporary digital communication and branding strategies.

KEYWORDS AI influencers, credibility, attractiveness, and content quality consumer trust, engagement, loyalty, Brand trust, SEM model.

Automatic Detection Thumbnail-Title Congruence in YouTube Branded Videos using Multimodal Machine Learning: Implications for View-based Engagement

르키아
순천향대학교 경영학과
–
최재원
순천향대학교 경영학과

This study investigates the impact of thumbnail-title emotional congruence on audience engagement in branded YouTube content, employing a hybrid machine-learning and statistical approach. A multimodal pipeline was designed by combining CLIP-based visual and textual embeddings, followed by 5 ML classifiers to predict the thumbnail-title congruence automatically. Among the classifiers (DT, ET, RF, XGBoost, and LGBM), LGBM achieved strong predictive performance: AUC = 0.82, AP = 0.87, on a data sample of 8082 thumbnail-title pairs labeled by human encoders. LGBM then automatically detected the probability of congruence of the full dataset (125,237 branded videos). These predicted probabilities were subsequently analyzed using a log-linear regression model, which revealed a significant positive relation between congruence and view-based engagement with $\beta = 1.44$, $p < .001$, $R^2 = 0.117$. A 0.10 increase in the predicted congruence corresponded to approximately a 15.5% rise in expected views, demonstrating that emotionally aligned thumbnails and titles enhance audience attention and video visibility, validating the effectiveness of the hybrid ML-statistical modeling framework.

KEYWORDS Thumbnail-title congruence, Multimodal ML, View-based engagement, YouTube branded videos, Hybrid ML-statistical model

Track
4-5

디지털 업무환경과 서비스

좌장: 박상철 대구대학교 교수

생성형 AI 시대 개발자의 과의존과 중독

안재영(연세대학교)
채송화(KT DS)
임혜정(연세대학교)
윤혜정(이화여자대학교)

디지털 인증 맥락에서 본 그림자노동의 구조적 전환: 행위자-네트워크 이론 관점의 분석

박상철 (대구대학교)
고준 (전남대학교)

호텔 종사원의 사회적 평판과 자기조절효능감이 서비스 대응성에 미치는 영향: 사회인지이론관점

권나현(경희대학교)
정남호(경희대학교)

생성형 AI 시대 개발자의 과의존과 중독

안재영
연세대학교
–
채송화
KT DS
–
임혜정
연세대학교
–
윤혜정
이화여자대학교

생성형 인공지능(Generative Artificial Intelligence; GAI)은 IT서비스 산업을 비롯하여 다양한 산업 전반의 생산성과 의사결정 구조를 혁신적으로 변화시키는 핵심 기술로 자리매김하고 있다. ChatGPT, Gemini, Copilot, CodeWhisperer 등과 같은 GAI 도구들은 코드 자동 완성, 디버깅, 문제 해결 지원 등 개발자의 주요 업무 과정을 자동화함으로써 업무 효율성과 정확성을 크게 향상시키고 있다. 이러한 기술적 편리함은 동시에 개발자들이 GAI의 결과물을 비판 없이 수용하거나 문제 해결의 상당 부분을 GAI에 의존하는 새로운 형태의 기술 과의존 현상을 초래하고 있다. 2023년 OECD 디지털 경제 보고서에 따르면, GAI의 확산은 혁신적 기회를 제공하는 동시에 데이터 편향, 신뢰성 저하, 비판적 사고력 약화 등의 부작용을 초래할 수 있음을 경고하였다. 반복적인 개발 업무가 자동화되면서 개발자들은 효율성의 이점을 얻는 한편, 독립적 사고력과 창의적 문제 해결 역량의 약화라는 역효과에 직면하고 있는 것이다. 그럼에도 불구하고 IT서비스 산업 내에서 개발자의 GAI 과의존을 실증적으로 분석한 연구는 여전히 부족한 실정이다. 기존의 연구들이 주로 교육이나 의료 등 특정 맥락에서의 윤리적, 신뢰 문제를 다뤘던 반면, 본 연구는 실제 산업 현장에서 활동하는 개발자들을 중심으로 생성형 AI 의존이 어떻게 형성되고, 어떤 심리적·행동적 결과가 중독으로 이어지는지를 규명하고자 한다.

본 연구의 주요 목적은 IT서비스 산업에서 개발자들이 GAI 사용 과정에서 과의존 상태로 발전하여 중독에까지 영향을 미치는 요인을 탐색하고, 그 인과구조를 규명하는 것이다. 4차 산업혁명 시대의 핵심 산업인 소프트웨어 분야는 인공지능, 빅데이터, 클라우드, 사물인터넷 등 첨단 기술의 융합을 통해 급속한 성장을 이루고 있으며, 이 중에서도 GAI는 개발자의 지식 탐색, 코드 작성, 테스트 자동화 등 전 과정에서 혁신을 주도하고 있다. GAI의 도구들은 실시간 코드 자동완성과 오류 수정 기능을 제공하여 개발 효율성을 극대화하고 있으며, 이러한 기술적 진보는 개발자의 업무 자동화와 사고 과정의 외주화를 가속화하고 있다. 하지만 GAI의 확산은 개발자의 인지적 자기결정력(self-determination)을 약화시키는 부작용을 수반한다. 과도한 GAI 의존은 개발자의 독립적 문제 해결 능력을 저하시킬 뿐 아니라, GAI가 제시하는 결과를 검증하지 않고 수용하는 '수동적 학습 습관'을 강화시켜 창의적 사고를 위축시킬 가능성이 있다.

‘과의존(Overdependence)’은 기술이나 도구에 대한 통제 상실, 내성, 금단 증상, 습관화된 사용 등의 특징을 지녔으며, 결국 중독으로 연결될 수 있는 개념으로 디지털 미디어 중독 연구에서 발전되어왔다. 스마트폰, 유튜브, SNS 등 과의존 현상에서 확인되듯, 사용자는 즉각적인 편의성과 보상에 끌려 점차 자기 통제력을 상실하고 판단 기능이 약화되는 경향을 보인다. 즉, GAI 과의존은 단순한 사용 빈도나 시간의 문제가 아니라 중독으로 연결되어 내성, 갈망, 강박적 사용 등 심리적·행동적 현상으로 이해된다. 최근 선행 연구에서도 GAI 과의존이 학습자 및 전문가의 창의력, 자기조절능력, 문제해결력 저하로 이어질 수 있음이 보고되었으나, 이러한 문제를 실제 산업 현장의 개발자를 대상으로 실증적으로 분석한 연구는 거의 이루어지지 않았다. 따라서 학문적 공백을 메우고, 개발자의 업무 환경 속에서 발생하는 GAI 중독의 심리적 기제와 결과를 실증적으로 규명한다는 점에서 의의가 있다.

이에 본 연구는 자기결정이론(Self-Determination Theory; SDT)을 이론적 기반으로 한다. SDT는 인간의 행동이 심리적 욕구의 충족 여부에 따라 내적·외적 동기로 구분된다고 설명한다. GAI 활용 맥락에서 (1) 개발자가 GAI에 의해 통제되지 않고 스스로 선택한다고 느끼는 정도’를 의미하며, (2) AI를 활용하여 문제를 해결할 수 있다는 자신감, (3) GAI 및 동료 개발자와의 상호작용에서 느끼는 사회적 유대감이다. 선행 연구에 따르면 충분히 충족되지 않을 때 개인의 기술 과의존 성향이 심화된다고 밝혀졌으며, 이러한 결과는 본 연구가 개발자의 GAI 과의존 요인을 분석하는 이론적 근거가 된다. 따라서 개발자의 내적 동기와 외적 동기 요인이 GAI 과의존에 미치는 영향을 분석하고, 과의존이 다시 GAI 중독으로 이어지는 인과 구조를 실증적으로 검증하였다.

이를 위해 연구 방법으로는 질적 연구와 양적 연구를 병행한 혼합 연구 방법론을 적용하였다. 1단계에서는 질적 연구로 IT서비스 산업 종사 개발자 13명을 대상으로 심층 인터뷰를 수행하였다. 인터뷰는 GAI의 사용 동기, 의존 수준, 문제 해결 방식의 변화 등을 중심으로 진행되었으며, 수집된 자료를 주제 분석(thematic analysis)으로 5개의 주요 요인(GAI 유용성, 지원 효과성, 개인 맞춤화, 시간 절약, 상대적 이점)을 도출하였다. 2단계에서는 질적 분석 결과를 토대로 자기결정이론을 이론적 틀로 설정하고, 연구모형을 구성하였다. 이후 온라인 설문조사를 실시하였으며, PLS-SEM을 활용하여 인과관계를 검증하였다. 그 결과, GAI 지원 효과성, 개인 맞춤화, 상대적 이점이 GAI 과의존에 유의한 영향을 미치는 것으로 나타났으며, GAI 과의존은 GAI 중독을 유의하게 예측하는 변수로 확인되었다.

이러한 결과는 GAI 기술이 개발자의 업무 효율성과 생산성을 향상시키는 도구로서 유용하지만, 동시에 자율성과 유능성의 결여를 초래하여 과의존으로 이어질 수 있음을 보여준다. 특히 GAI 지원 효과성과 개인 맞춤화 기능은 사용자의 편리성을 높이지만, 개발자가 스스로 사고하고 판단하는 과정을 대체함으로써 인지적 자기결정력을 약화시키는 요인으로 작용한다. 따라서 개발자의 자기결정 능력과 비판적 사고를 유지·강화할 수 있는 관리 전략을 병행할 필요가 있다.

디지털 인증 맥락에서 본 그림자노동의 구조적 전환: 행위자-네트워크 이론 관점의 분석

박상철
대구대학교 경영학부
–
고 준
전남대학교 경영학부

디지털 기술의 발전은 사용자에게 편의성을 제공하는 동시에, 시스템 운영을 위해 사용자가 수행해야 하는 보이지 않는 노동을 확대시켜 왔다. 본 연구는 이러한 현상을 '그림자노동(shadow work)'의 관점에서 접근하여, 디지털 인증 절차의 변화 속에서 사용자 노동의 구조적 전환을 탐구하였다. 본 연구는 2020년과 2025년에 두 대학의 경영정보시스템 교과목에서 수집된 학생 보고서를 질적으로 분석하여, 과거의 절차 수행 중심 그림자노동과 현재의 데이터 제공 중심 그림자노동을 비교하였다. 행위자-네트워크 이론을 분석 틀로 활용하여, 인간(사용자)과 비인간(기술)이 신뢰를 공동 생산하는 과정 속에서 그림자노동의 위치와 성격이 어떻게 변모했는지를 규명하였다. 분석 결과, 디지털 인증의 자동화는 노동의 소멸이 아니라 형태의 재배치를 의미하며, 사용자는 절차 수행자에서 기술 운영의 협력적 행위자로 전환되고 있었다. 이는 그림자노동이 단순히 불편함의 총량이 아닌, 인간과 기술이 함께 유지하는 관계적 협력 구조로 변화했음을 보여준다. 본 연구는 이러한 결과를 바탕으로, 디지털 인증 연구에서 사용자 경험(UX)과 기술적 신뢰의 관계를 재해석하고, 향후 AI 기반 인증 환경에서 나타날 새로운 그림자노동의 가능성을 논의하였다.

2025 한국지식경영학회
추계학술대회

호텔 종사원의 사회적 평판과 자기조절효능감이 서비스대응성에 미치는 영향

: 사회인지이론 관점

권나현, 정남호
경희대학교 스마트관광원

2025 한국지식경영학회
추계학술대회

Contents

01. Introduction
02. Literature Review and Hypotheses Development
03. Methodology
04. Results
05. Conclusion and Discussions

1 Introduction

호텔산업에서 서비스품질

- 고객들의 만족과 재방문을 유도하기 위해서는 서비스 품질이 중요하기 때문에, 서비스 품질 전략이 중요하다 (Othman et al. 2019)
- 호텔 사업의 성공 또는 실패는 고객과 직원 간의 협업에의해 결정된다(Abdullah, 2018)

호텔산업에서 인적자원관리

- 호텔 고객들의 니즈가 다양해지면서 고객접점 서비스를 제공하는 종사원의 태도는 서비스 품질에 중요한 영향을 미친다 (김진옥·한주형, 2021)
- 객실, 식음, 각종 부대시설 등 다양한 부서에서 근무하는 직원들의 혁신행동은 고객의 경험과 서비스 만족에 중요한 요소이다 (Dhar, 2015, Lim et al., 2024).
- 고객의 관점에서 바라보는 직원들의 태도가 고객의 니즈를 빠르게 파악하고, 더 나은 서비스를 제공하 하는 의지가 강하다(Homburg et al.,2009; Huo et al., 2019).
- 직원들이 회사의 지원을 인식할 때, 업무에 더 몰두하고, 자발적인 행동을 한다

자발적인 참여

- 호텔산업에서 경쟁 우위를 제공하기 위해서는 자발적인 참여가 필수적이다 (Yanet al., 2023).
  (Cheng et al., 2023; Luu, 2024)).
- 자발적인 행동을 보이는 직원은 고객의 기대를 충족시키고, 동료의 작업을 효율적으로 수행하도록 도와 서비스 품질 향상에 도움을 준다

1 Introduction

Research gap

- 리더십, 심리적 임파워먼트·성취욕구 등 개인의 내적 요인 -> 비보상적 행동 (윤대혁, 정순태, 2006; 강영순, 김정훈,2013; 권인수, 최영근,2011)
- 조직문화·공정성 등의 조직적 요인이 -> 비보상적 행동 (강대석, 차동옥, 2004;, 조윤형, 최우재,2011;

→ 외적 사회적 요인과 내적 개인 요인을 통합적으로 고려 연구 미흡

Research purpose

- 호텔종사원의 사회적인지와 개인의 자기조절 신념이 조직 내 자발적 행동에 미치는 영향을
  사회인지이론(Social Cognitivie Theroy) 을 적용하여 실증분석
- 조직시민행동을 매개로 서비스 대응성에 미치는 영향을 살펴보고자 한다

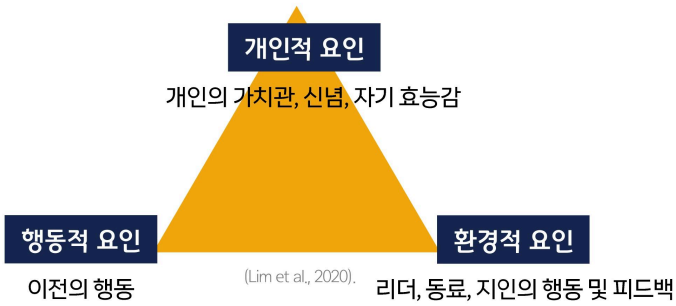


2 Literature Review and Hypotheses Development

2.1 사회인지이론(Social Cognitive Theory)

- 사회 학습 이론의 발전된 형태로서 개인의 행동을 이해하기 위해 사용되는 이론
- 인간의 행동을 결정할 때에는 개인적, 환경적, 행동적 요소들이 상호 작용(삼원 상호결정성)하여 영향 미침 (Bandura, 1984, 1997, 2001).

서로 다른 원천이 동일한 정도를 갖지 않으며, 상호 영향이 모두 특정 시점에 동시에 발생하는 것은 아니다 (Bandura, 1991).



2 Literature Review and Hypotheses Development

2.2 사회적평판

- 본인의 직업이 사회적으로 얼마나 자랑스러운지를 인식하는 정도
  => 개인의 내적 자긍심이 아니라, 사회가 부여한 가치를 개인이 지각한 요인 (고동우, 전병길,2002)

직업존중감 • 자신이 현재 종사하거나 미래에 종사하게 될 직업에 대하여 판단하는 가치평가 및 기대의 정도를 직업 존중감 (Miller, 1999).

하위 요인 사회적평판 | 부정적 대우 | 상대적 효익 (고동우, 전병길,2002)

기존 문헌

- Miller(1999) 직업존중감의 최종 문항에서 제외된
  '직업위신(occupational prestige)'과 비슷한 개념 (Hope, 1982)
  : 객관적으로 고려되기보다 문화적으로 결정된 방식으로 인식 => 삭제

본 논문

- 사회적, 문화적 영향(집단주의 성향) : 다른사람과 연결
- 직업의 위신, 타인의 시선, 체면문화 (유교문화)
  => 사회의 평가 (환경적 요인)
  (Fung, Ladkin, & Osman, 2022)

2 Literature Review and Hypotheses Development

2.3 자기조절 효능감 (Self- regulatory efficacy)

- 개인이 목표 달성을 위해 자신의 생각·감정·행동을 스스로 모니터링하고, 규제하고, 지속할 수 있다고 믿는 신념
- 과정의 통제 (주의 집중, 충동 억제, 노력 유지, 전략 전환)에 초점을 둔 개념 (Bandura, 1986, ; Zimmerman,1992)

자기효능감 • 개인이 다양한 상황에서 과업을 수행하는 자신의 기술에 대한 직원의 믿음 (Bandura, 1991, 1997)



2 Literature Review and Hypotheses Development

2.4 조직시민행동

- 보상을 바라지 않고 자의에 의해서 동료들을 도와주고 필요한 정보를 공유하며 조직의 공동의 이익을 위해서 협력하는 행동 Organ(1988)
- 조직의 보상시스템으로부터 영향을 받는 것이 아닌, 구성원 개인의 자발적인 행동

호텔서비스 환경에서는 고객과의 모든 접촉상황을 관찰하기 어렵기 때문에 고객들의 자발적 행동이 중요함 (안관영, 곽영환 , 2003)

조직적 차원	성실성	조직에서 요구하는 최소 수준이상의 근태율 혹은 규칙을 준수하는 것으로, 야근을 하거나 시간을 준수하는 등의 행동
	시민정신	조직의 성과를 높이기 위해서 조직에 관심을 가지고 책임감 있게 보이는 행동 (중요 회의 참석, 공지 확인 등)
개인적 차원	예의성	업무와 관련하여 어떤 결정을 내렸을 때, 발생할 수 있는 문제들을 사전에 파악하고 동료들에게 공유 하는 등의 노력
	이타성	조직 내부에서 업무와 관련하여 상사 혹은 동료를 돕는 행동

(Williams & Anderson, 1991),

2 Literature Review and Hypotheses Development

0 사회적평판 (외적요인) -> 조직시민행동

- 동양문화권인 한국에서 사회적 인식은 개인에게 중요한 요소이며, 자신의 업무수행에 노출되는 호텔직원들의 직무만족에 영향을 미친다 (이상희, 2015; 문상정, 김기성 2017)
- 사회적 평가에 대한 민감성 혹은 인식이 개인의 직무태도에 영향을 미친다 (장해수 외, 2008; 김민주, 2011; 문상정·김기성, 2017)
- 높은 사회적 평가는 직원의 직무만족에 영향을 미치며, 조직 내 업무에 투입하기 위해 노력한다 (홍지웅·홍아정, 2021)
- 긍정적인 사회적 평가 인지하는 직원은 높은 직무열의와 타인에 대해서 최선을 다기 위해 노력한다 (송인숙 권상집, 2022)

Hypothesis 1	H1a. 사회적 평판이 성실성에 정(+)의 영향을 미칠 것이다. H1b 사회적 평판이 시민정신에 정(+)의 영향을 미칠 것이다. H1c 사회적 평판이 예의성에 정(+)의 영향을 미칠 것이다. H1d. 사회적 평판이 이타성에 정(+)의 영향을 미칠 것이다.

2 Literature Review and Hypotheses Development

0 자기조절효능감 -> 조직시민행동

- 자기효능감이 높은 사람들은 스스로 스트레스 요인을 찾고, 줄이려 노력하며 서비스 지향적인 태도를 가진다 (Leiter, 1991, 이형룡 · 이승호 · 심지연, 2013)
- 높은 자기효능감은 직무 만족에 긍정적 영향과 높은 직무성과로 이어져 고객 만족 및 기업 수익에 긍정적 영향을 미쳤다(서창적 · 김효정 · 이호택 · 장하다, 2010).
- 직무 만족이 높을수록 조직을 위한 자발적인 행동을 하려는 의도가 증가한다. (박영국, 2009)

Hypothesis 2	H2a 자기조절효능감은 성실성에 정(+)의 영향을 미칠 것이다. H2b 자기조절효능감은 시민정신에 정(+)의 영향을 미칠 것이다. H2c 자기조절효능감은 예의성에 정(+)의 영향을 미칠 것이다. H2d 자기조절효능감은 이타성에 정(+)의 영향을 미칠 것이다.

2 Literature Review and Hypotheses Development

2.5 서비스 대응성

- 고객을 돕고자 하는 의지와 즉각적인 서비스를 제공하고자 하는 의지 (A.Parasuraman, 1988)
- 고객의 요청, 질문 및 불만 사항 처리에 주의와 정확성을 강조 (A.Parasuraman et al,, 2013)

서비스의 특징
- 서비스는 상품과 다르게 무형적이고, 사물이 아닌 과정 (Abdullah & Rahman, 2015)
- 서비스는 보이지 않기 때문에 공급자가 밝히고, 소비자가 측정하기 때문에 파악하기 어려움 (Ali et al. 2021)

서비스품질
- 고객에게 제공되는 서비스의 수준이 그들의 기대에 부합하는 정도를 측정하는 것 (Lewis & Booms, 1983)
- 서비스 공급자에 대한 고객의 기대와 서비스에 대한 평가 간의 차이Abdullah & Othman, 2021

호텔의 서비스 품질은 손님의 기대에 부응하는 서비스를 제공하는 것, 경쟁우의를 유지하는 전략이다 (Top & Ali, 2021).

하위 요인 유형성 | 신뢰성 | **대응성** | 보증성 | 감정이입

2 Literature Review and Hypotheses Development

0 조직시민행동 -> 서비스 대응성

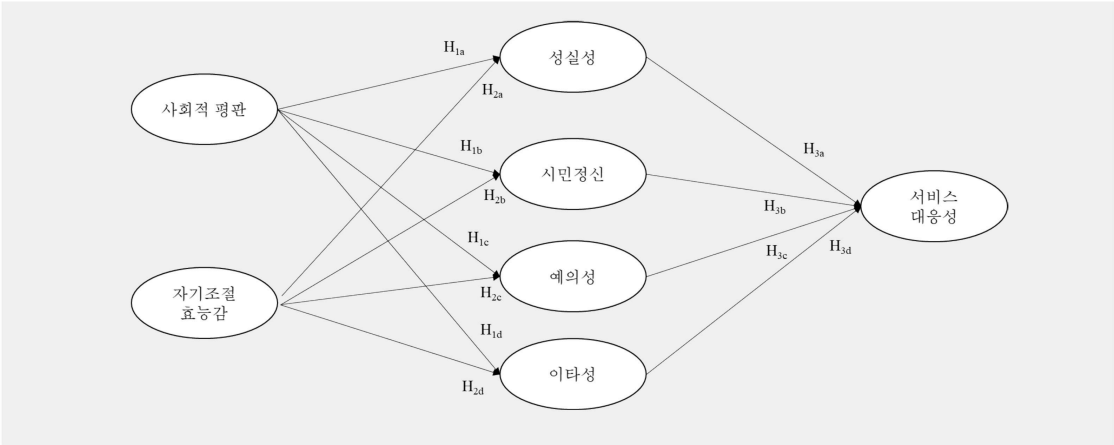
- 종업원들의 자발적인 행동은 조직성과에 긍정적인 영향을 미친다 (Podskoff et al., 1990)
- 호텔 종사원들의 자발적인 행동은 서비스 품질에 긍정적인 영향을 미쳤다 (안관영, 곽영환 2003)

Hypothesis 3

H3a 성실성은 서비스 대응성에 정(+)의 영향을 미칠 것이다.
H3b 시민정신은 서비스 대응성에 정(+)의 영향을 미칠 것이다.
H3c 예의성은 서비스 대응성에 정(+)의 영향을 미칠 것이다.
H3d 이타성은 서비스 대응성에 정(+)의 영향을 미칠 것이다.

3 Methodology

Research Model



3 Methodology

구 분		내 용		
모집단		국내 5성급 및 4성급 호텔 종사자		
표본 추출 방법		확률적 표본 추출 방법 (단순 무작위 추출 방법)		
설문지 배포 및 회수		온라인	오프라인	총
	회수된 설문지	80	152	232
	제외된 설문지	8	12	20
	분석에 사용된 유효 표본			212

설문 대상 호텔

■ 5성급호텔 (여의도 콘래드 서울 호텔, 앰배서더 서울 풀만 호텔, 더 플라자 호텔, 더 웨스틴 조선 서울, JW메리어트 호텔 서울, 그랜드 하얏트 서울, 그랜드 머큐어 앰배서더 호텔 앤 레지던스 서울 용산, 노보텔 앰배서더 서울 동대문, 노보텔 앰배서더 서울 용산, 몬드리안 서울 이태원 (구. 호텔캐피탈), 파라스파라 프라나 오너스)

■ 4성급 호텔 (그래비티 서울 판교 오토그래프 컬렉션, 라이즈 오토그래프 컬렉션 바이 메리어트, 서울가든호텔, 코트야드 바이 메리어트 서울 타임스퀘어, 켄싱턴 호텔 여의도, 레스케이프 호텔)

3 Methodology

Analysis tool
- Smart PLS4.0
- SPSS 22

Methodoldy
- Frequency Analysis
- Confirmatory Factor Analysis(CFA)
- Reliability Testing
- Structural Equation Modeling(SEM)

- Factor loading > 0.7
- Cronbach's a > 0.7
- CR > 0.7
- AVE > 0.5

구분		빈도 (명)	비율 (%)	구분		빈도 (명)	비율 (%)
성별	여자	101	47.9	연령	20대	77	36.5
					30대	78	37.0
	남자	110	52.1		40대	43	20.4
					50대이상	13	6.2
	합계	211	100		합계	211	100
학력	고졸 이하	6	2.8	직급	사원급	87	41.2
	2년제졸업/재학	45	21.3		주임급~대리급	63	29.9
	4년제졸업/ 재학	121	57.3		과장급~차장급	48	22.7
	대학원졸업/ 재학	39	18.5		임원이상	13	6.2
	합계	211	100		합계	211	100
전공 일치 여부	전공일치	154	73.0	고용 형태	비정규직	108	51.2
	전공 불일치	57	27.0		정규직	103	48.8
	합계	211	100		합계	211	100
부서	객실 부문	101	47.9	근속 년수	1년미만	64	30.3
	식음료 부문	45	21.3		1년이상~3년미만	40	19.0
	조리 부문	7	3.3		3년이상~6년미만	44	20.9
	마케팅 부문	14	6.6		6년이상~9년미만	22	10.4
	관리부문 (백오피스)	32	15.2		9년이상~12년미만	13	6.2
	기타	12	5.7		12년이상	28	13.3
	합계	211	100		합계	211	100

4 Results

요인명	연구변수 및 항목	표준요인 부하량	C.R	AVE	Cronbach's α
직업 평판	나는 내 직업으로 인해 사회적으로 인정받을 수 있다.	0.814	0.899	0.692	0.853
	내 직업은 여러 가지 면에서 좋은 직업이다.	0.789			
	내 직업은 자랑할 만 하다.	0.828			
	가족들은 내 직업에 대해 자랑스럽게 생각한다.	0.662			
	내 직업은 이 사회에 공헌하는 게 많다.*	-			
자기조절효능감	어떤 문제에 대한 나의 판단은 대체로 정확하다*	-	0.846	0.647	0.715
	나는 어떤 일의 원인과 결과를 잘 분석해 낼 수 있다.	0.646			
	나는 일이 잘못되고 있다고 생각되면 빨리 바로 잡는다.	0.644			
	나는 내가 할 수 있는 일과 그렇지 않은 일을 판단할 수 있다.	0.736			
	나는 항상 목표를 세우고 그것에 비추어서 일의 진행 상태를 확인한다.*	-			
성실성	나의 출근율은 표준수준보다 높다.*	-	0.838	0.634	0.777
	공식적으로 주어진 휴식외에 추가로 휴식을 가지지 않는다.*	-			
	아무도 보지않을 때에도 회사 규칙과 규정을 준수한다.	0.668			
	나는 가장 양심적인 직원중에 한명이다.	0.81			
	정직하게 하루를 일해야 정직한 댓가가 있다고 믿는다.	0.749			

4 Results

시민정신	필수는 아니지만, 중요하다고 생각되는 회의에 참석한다.	0.621	0.817	0.530	0.757
	필수는 아니지만, 회사의 이미지에 도움을 주는 행사에 참석한다.	0.679			
	조직의 변화에 뒤처지지 않는다.	0.762			
	회사 내 공고나 안내를 지속적으로 파악한다.	0.585			
예의성	동료들과의 마찰을 방지하기 위해서 조치를 취한다.	0.734	0.890	0.672	0.778
	나의 행동이 다른 동료들의 업무에 어떻게 영향을 줄지 고려한다.	0.791			
	타인의 권리를 남용하지 않는다.	0.557			
	동료에게 문제를 일으키는 것을 피하려고 노력한다.*				
	나의 행동이 동료에게 어떻게 영향이 있을지 고려한다.	0.662			
이타성	결근한 동료를 돕는다.*	-	0.945	0.812	0.881
	업무가 과중한 동료를 돕는다.	0.753			
	요청받지 않았지만, 신입사원을 돕는다.	0.782			
	업무에 관해 문제가 있는 동료를 기꺼이 돕는다.	0.869			
	항상 주변에 도움의 손길을 줄 준비가 되어있다.	0.825			
서비스 대응성	고객들에게 정해진 시간 안에 서비스 약속을 지킨다.*	-	0.946	0.855	0.886
	고객이 문제가 있을 때는 공감해주며 안심을 시켜준다.	0.861			
	고객들에게 신뢰를 준다.	0.908			
	고객들에게 정확한 시간에 서비스를 제공한다.	0.786			
	서비 스제공과 관련된 고객에 대한 정보를 정확하게 기록한다.*	-			

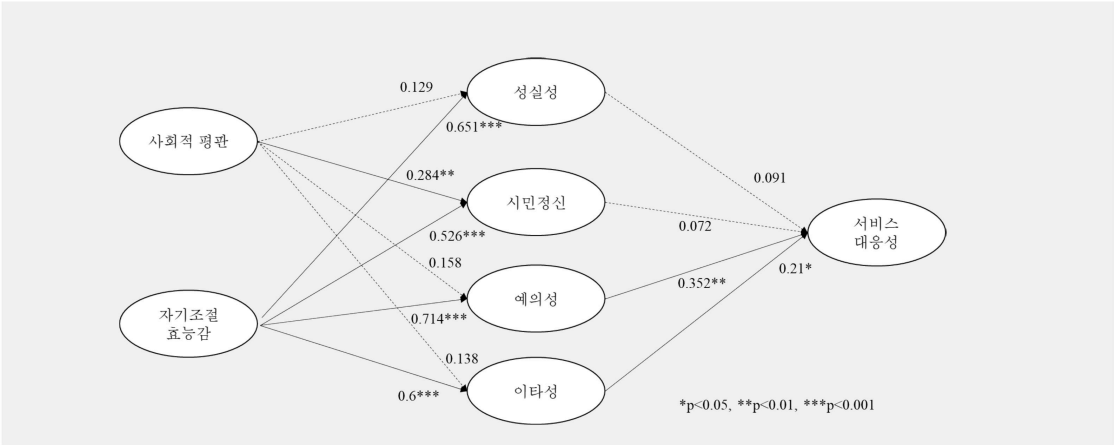
4 Results

<표 3> 측정개념의 평균과 표준편차간 상관관계

측정개념	평균	표준편차	상관관계					
			(1)	(2)	(3)	(4)	(5)	(6)
(1) 직업 평판	3.648	0.679	**0.832**					
(2) 자기조절효능감	3.813	0.543	0.317**	**0.805**				
(3) 성실성	3.898	0.711	0.332**	0.463**	**0.796**			
(4) 시민정신	3.577	0.633	0.388**	0.349**	0.390**	**0.728**		
(5) 예의성	4.046	0.524	0.391**	0.450**	0.472**	0.472**	**0.820**	
(6) 이타성	4.110	0.565	0.347**	0.418**	0.427**	0.343**	0.533**	**0.901**
(7) 서비스 대응성	4.130	0.606	0.313**	0.424**	0.376**	0.347**	0.472**	0.434**

3 Methodology

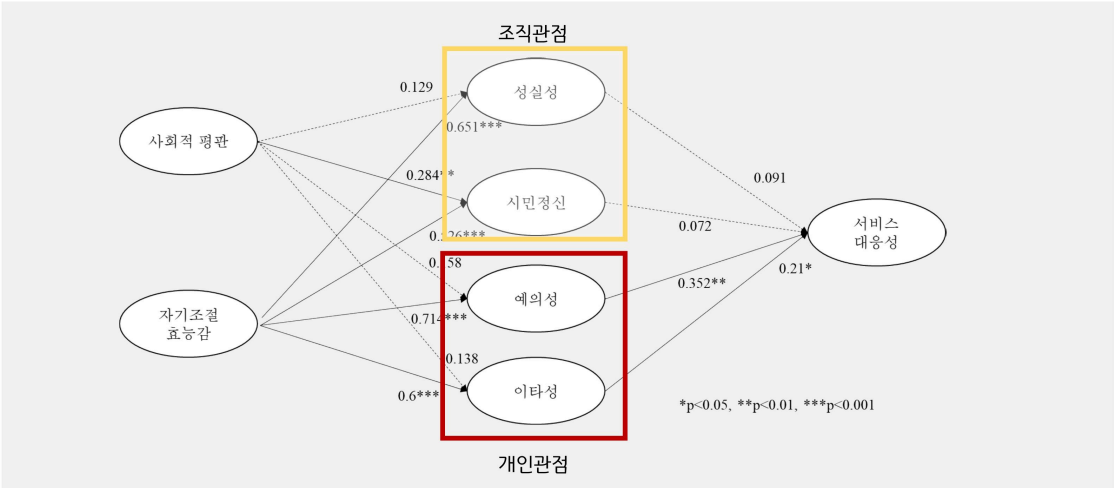
Research Model



3 Methodology

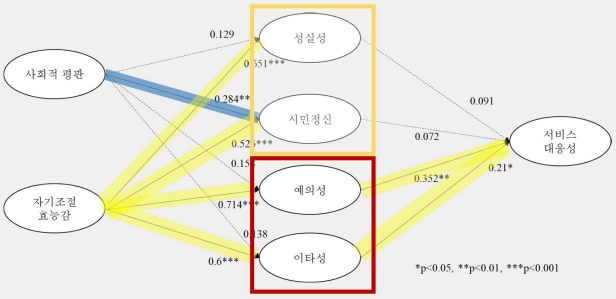
19/ 24

Research Model



4 Results

가설검증 결과



χ^2 = 380.850, df = 262, χ^2/df = 1.454, p = 0.000,
GFI = 0.874, AGFI = 0.844,
NFI = 0.854, CFI = 0.948, RMSEA = 0.047

가설	경로			표준화 경로계수	p값	가설검증
H_{1a}	직업평판	→	성실성	0.129	0.142	기각
H_{1b}	직업평판	→	시민정신	0.284	0.002	채택
H_{1c}	직업평판	→	예의성	0.158	0.061	기각
H_{1d}	직업평판	→	이타성	0.138	0.095	기각
H_{2a}	자기조절효능감	→	성실성	0.651	***	채택
H_{2b}	자기조절효능감	→	시민정신	0.526	***	채택
H_{2c}	자기조절효능감	→	예의성	0.714	***	채택
H_{2d}	자기조절효능감	→	이타성	0.600	***	채택
H_{3a}	성실성	→	서비스대응성	0.091	0.353	기각
H_{3b}	시민정신	→	서비스대응성	0.072	0.450	기각
H_{3c}	예의성	→	서비스대응성	0.352	0.001	채택
H_{3d}	이타성	→	서비스대응성	0.210	0.017	채택

5 Conclusion and Discussions

Conclusion

√ 사회적 평판 → 조직시민힝동에 영향을 미친 요인

- 사회적 평판 → 시민정신 유의
  - 개인 수행 중심의 행동(예의성·이타성)
  - 조직의 이미지·명예를 고려한 조직참여형 행동에 영향을 줌

√ 자기조절효능감 → 조직시민행동에 영향을 미친 요인

- 자기조절효능감 → 모든 조직시민행동(OCB) 하위요인(성실성·시민정신·예의성·이타성) 유의한 긍정적 영향
- 개인의 자기통제·감정조절 신념 → 자발적 행동의 핵심 동기임을 입증

√ 예의성, 이타성 → 서비스 대응성(품질)에 미치는 영향

- 고객 중심의 공감적·자발적 행동 → 서비스 품질에 영향

5 Conclusion and Discussions

이론적 시사점

√ OCB의 조직차원, 개인차원을 실증적으로 확인

사회적 평판 → 조직시민행동 (조직지향: 시민정신·성실성)
자기조절효능감 → 조직시민행동 (개인지향: 예의성·이타성)
특히 서비스 맥락에서 조직지향 행동은 간접효과, 대인지향 행동은 직접 서비스 품질에 영향을 미침

- 외적 요인(사회적 평판)과 내적 요인(자기조절효능감)이서로 다른 형태의 자발적 행동(OCB)을 유발함을 보여줌.

√ SCT의 '환경-개인-행동' 구조 실증

- 외적 요인(사회적 평판)과 내적 요인(자기조절효능감)이서로 다른 형태의 자발적 행동(OCB)을 유발함을 보여줌

5 Conclusion and Discussions

실무적 시사점

√ 사회적평판 제고 = 시민정신 강화 (조직 몰입 향상)

내부 브랜딩·직업 가치 홍보로 "우리 직업이 사회적으로 자랑스럽다"는 인식 강화 →직원의 참여·제안·조직 관심 행동 촉진.

√자기조절역량 개발 = 서비스 품질 강화

감정노동 상황에서 자기통제·감정조절 훈련 프로그램 (시나리오 코칭, 피드백 루틴) 도입 →공감적 대응·도움 행동 향상

√조직문화 개선

관리자 평가체계에 자발적 행동과 협력행동의 인정을 반영 → 직원이 '보상 없이도 행동할 이유'를 내면적으로 강화.

√분리된 관리 전략

조직지향 행동(시민정신·성실성) 은 조직 가치·브랜딩 강화로, 대인지향 행동(예의성·이타성) 은 감정조절·공감훈련 중심으로 육성.

	Memo	

	Memo	

	Memo	

	Memo	

2025 공동 학술대회

Sponsors' Advertisement

Barun ICT Research Conference
한국지식경영학회 추계학술대회
ISACA Korea Conference





AI는 언제나
사람을 향해야 하기에
기술보다 안심을 먼저 생각합니다

기술이 앞서가도
불안은 뒤에 남지 않도록
사람의 편에서 믿고 함께하는 AI

유플러스가 익시 가디언으로
모두가 안심할 수 있는
AI 시대를 시작하겠습니다

사람 중심. 안심 지능.

Assured Intelligence

LG U+

익시 가디언 ixi-Guardian
LG유플러스만의 차별적 AI 보안 기술 브랜드

Deep Fake 탐지 기술로 위·변조 음성을 판별하는 **Anti-Deep Voice**
LG AI 연구원의 초거대 AI '**EXAONE**' 기반으로 서버 통신 없이 안전하게 스마트폰 안에서 구동 되는 **On-Device AI**
국방용 보안 인증 수준의 세계적인 PUF(물리적 복제 방지) 기술을 독점 제휴한 **양자 보안(PQC)** '23년 7월~'27년 12월

우리엔아이에스 회사소개

WOORINIS
(주)우리엔아이에스

회 사 명	㈜우리엔아이에스	대 표 이 사	백창준
설 립 일	1999년 4월 12일	보 유 자 격	정보통신공사업, 소프트웨어사업자등록 등
사 업 내 용	유,무선 네트워크 구축 & 네트워크 공사 보안 & 솔루션 제안 및 구축 IT 인프라 시스템 설계 및 구축 통합 유지보수	직 원 수	48명
		홈 페 이 지	http://www.woorinis.com/
주 소	서울시 송파구 법원로 11길7, C동 1007호~1009호 우리엔아이에스 (문정동, 현대지식산업센터)	제 품 문 의	Sales@woorinis.com 02-3012-7800

우리엔아이에스 주요 사업 내용

1. 네트워크 & 네트워크 공사
Network & Network Construction

- 유,무선 네트워크 (cisco)
- 무선 네트워크 (HPE aruba networking)
- 네트워크 공사

2. 보안 & 솔루션
Security & Solution

- 차세대 방화벽 (paloalto)
- 시스템 접근제어 & 패스워드 관리 (Suhyunsoft)
- Task QoS (nstek)
- L4-L7 스위치 (f5)

3. 시스템
System

- HCI (NUTANIX)
- 가상화 시스템 (vmware)

4. 통합 유지보수
Integrated Maintenance

- 학교
- 기업
- 기관

SETA
SEEDGEN Education Training and Awareness

보안은 ON
인식은 SET
교육은 A급이니까!

SETA 정보보호 교육
기업 맞춤형 사이트 개설
및 임직원 맞춤형 교육 제공

악성메일 모의훈련
악성코드 이메일 대처를 위한
반복적인 모의훈련 실시

정보보호 캠페인
임직원 흥미를 유도하는
다양한 정보보호 캠페인 시행

사내 보안수준 확인
지속적인 보안수준 관리로
사내 보안문화 형성



세상에서 가장 좋은 **타**이밍을 마주하다,
Seedgen **E**ducation **T**raining and **A**wareness.

onSETA 바로가기

씨에이에스(C·A·S)는 정보통신기술(ICT)분야에서 글로벌 경쟁력을 갖춘

Consulting **Assurance** **Solution** 전문기업 입니다.

Global Leader as IT Assurance Services Group

감리·감사 및 보증



- 씨에이에스는 1995년 국내 최초 설립된 정보시스템 감리법인
- '감리 그 이상의 서비스'를 제공하는 ICT Audit/Assurance 서비스 분야 글로벌 리더 그룹
- 고객의 업무연속성 확보를 위한 솔루션 전문기업

IT 컨설팅 및 PMO



- 고객의 디지털 전환(DX)지향 지속발전가능체제를 위한 정보전략 컨설팅
- 고객의 디지털서비스플랫폼 기획·개발·운영을 위한 종합 컨설팅 서비스
- 전사 PMO방법론에 기반한 ICT개발 프로젝트의 성공보증 서비스 전문 기업

통합 정보보호



- 정부로부터 인가된 전문성과 독립성을 갖춘 정보보호전문서비스 기업
- 개인정보영향평가기관으로서 국내최대 실적의 개인정보 통합서비스 전문 그룹
- 보안인증 및 기반시설 취약점분석 및 모의해킹, 통합 정보보호 서비스 기업

ESGRC S/W



- AI형 감사정보시스템, 준법 관리, 리스크 관리, 보안포털, 컴플라이언스 솔루션
- 국내최초 내부회계관리제도 솔루션 상용화를 통한 내부통제감사 솔루션 제공
- ESGRC(환경, 사회, 거버넌스, 리스크 관리, 컴플라이언스)글로벌 전문 솔루션 기업

데이터 품질



- 데이터 산업법 제20조에 근거한 데이터품질인증기관(과기정통부 지정 제1호)
- 데이터 품질진단, 품질인증 사전컨설팅, 데이터 품질인증 등 다수의 데이터품질사업 수행 실적
- 데이터품질심사원, 데이터거래사, DBA/DAP/DQC 등 국내 최대 규모의 데이터 전문가 그룹

IT보안기술 및 위험관리 연구



- 감리방법론/평가방법론 연구개발
- 품질관리 방법론 연구개발
- 위험관리방법론 연구개발
- POCA 자동화 방법론 연구개발

▸ CAS 홈페이지 www.casit.co.kr
▸ 데이터품질인증 홈페이지 test.casit.co.kr

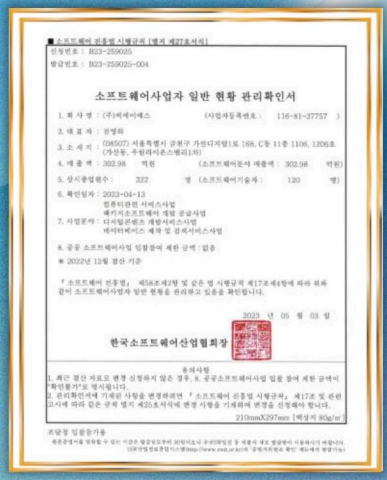
(주)씨에이에스(C·A·S)
Computer Assurance Services

지속발전가능 보증으로 고객의 가치극대화를 우선시하는

한국 컴퓨터보증서비스의 역사! ICT 전문기업 씨에이에스







고객의 문제해결과 새로운 가치창출을 위해 전문가 책임을 다하는 기업,
씨에이에스가 함께 합니다 !

Global Leader as IT Assurance Services Group

현재
- 정보보호 유공 장관표창(2024)
- 과학기술정보통신부 데이터 품질인증기관 지정(2023)
- 신한은행 감사패 수상(2024)
- 한국인터넷진흥원, 일자리 창출콘테스트 대상(2023)



2015 ~
- 경제부총리, 국가통계발전 기여 표창(2016)
- 기술혁신 중소기업(INNO-BIZ) 인증(2016)
- 과학기술정보통신부 정보보호 전문서비스 기업지정(2017)
- 한국데이터산업진흥원 데이터품질 인증심사기관 지정(2018)

2020
- KB국민은행 감사패 수여
- 선박해양플랜트연구소 공로패 수여
- 개인정보보호위원회 표창장(2021)
- 기업신용평가등급 A+획득(2022)

2010 ~
- 관세청, 광양세관 감시종합 정보시스템 구축(2011)
- 행정안전부 개인정보 영향 평가 기관 지정(2011)

2005 ~
- KISA 정보보호 관리 체계 (ISMS) 인증 획득(2007)

1995
- 국내 최초 IT 보증 서비스 회사 설립

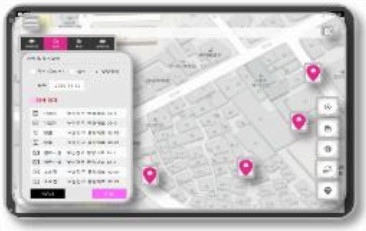
▸ CAS 홈페이지 www.casit.co.kr
▸ 데이터품질인증 홈페이지 test.casit.co.kr

(주)씨에이에스(C·A·S)
Computer Assurance Services

꿈 · 첨단 · 창의의 빛으로

대한민국 ICT를 이끌어가는 주식회사 투비콤





▪ 시스템 통합관리

- 상하수도 통합, 상수관망 최적 관리
- 전사자원통합관리, 자산관리정보시스템
- GIS기반 공공행정, 모바일 시설물관리





▪ 스마트 빌리지

- 스마트 영농지원, 스마트 생활편의
- 스마트 생활안전, 스마트 주민생활
- 스마트 빌리지 플랫폼

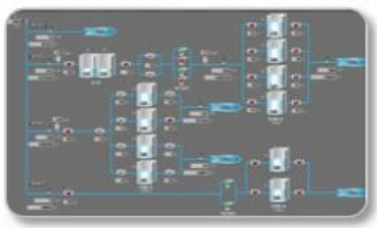




▪ 공공 SI

- 차세대 Census 통합관리시스템
- 외국인 행정정보 종합플랫폼
- 정보시스템 통합 유지보수
- 사회보장정보 통합플랫폼 유지관리





▪ 계측제어 및 SCADA 시스템

- 상하수도 · 환경기초시설 자동제어 시스템
- 원격검침시스템
- 스마트 급전제어장치



㈜투비콤

본사 : 서울시 금천구 가산디지털1로2 Tel. 02)2027-2041
우림라이온스밸리 2차 1119호 Fax. 02)2027-2042

고객의 미래가치를 실현하는 **AX전문기업**

세림TSG

Selim Technology Service Group



Realize AX for the future

AX

심층적 업무이해와 AI Agent, RAG 등
혁신기술을 접목해 AI 대전환을 선도합니다.

#범정부 초거대 AI　#국민비서 챗봇　#지능형 업무관리

Cloud

클라우드 도입을 위한 컨설팅부터 마이그레이션, 효율적 운영까지
End-to-End 서비스를 제공합니다.

#G-Cloud 운영　#Cloud Native　#MSP

Data Center

국내 최대 공공데이터센터의 경험과 노하우로 안전하고 효율적인
인프라를 설계, 구축, 운영, 유지관리 합니다.

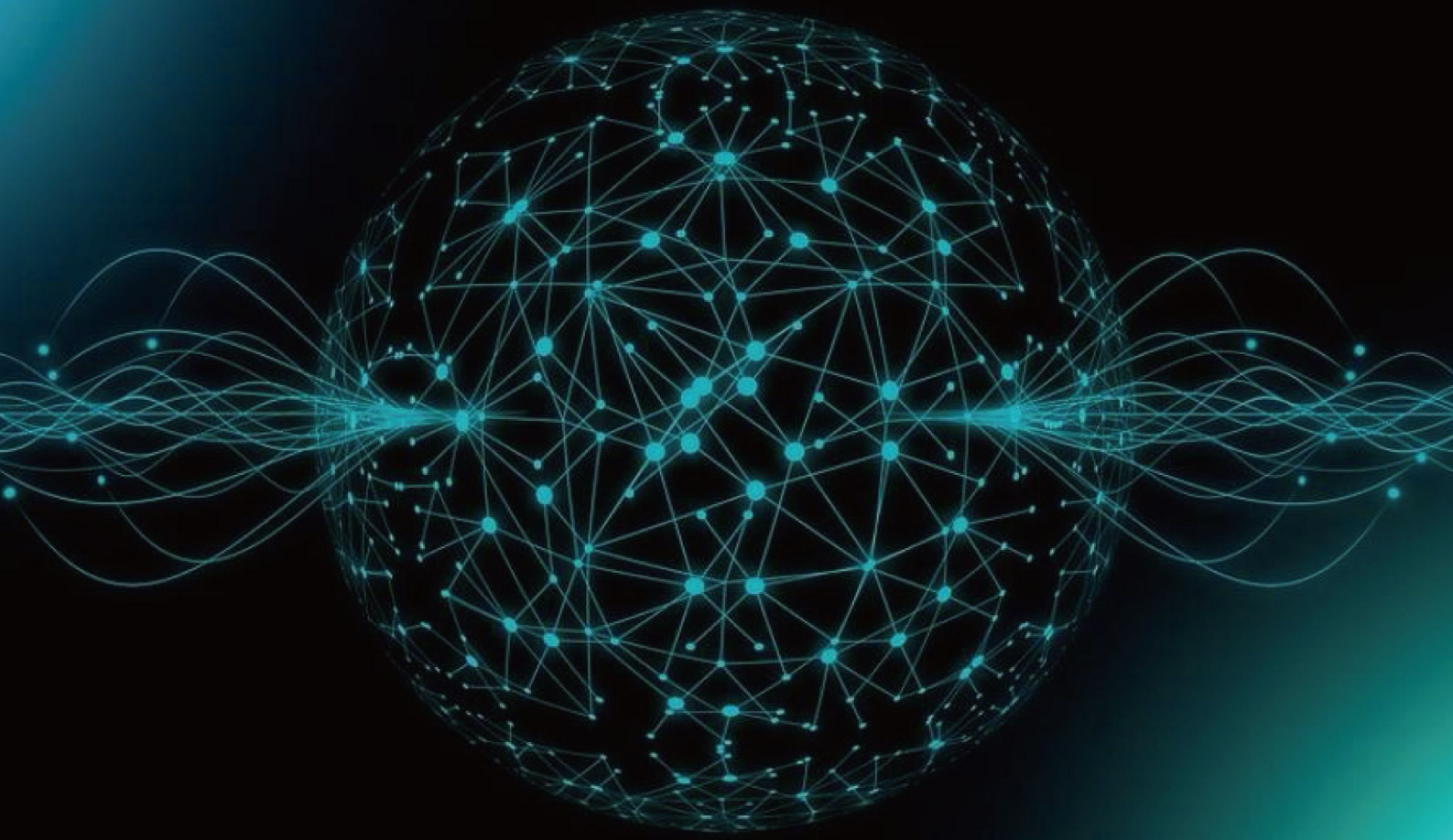
#통합운영　#정보자원 통합구축　#SDDC 아키텍처

본사 | 대전광역시 유성구 테크노1로 62-16 (관평동 755)
대표전화 | 042.488.7900　**팩스** | 042.488.7799　**이메일** | selimtsg@selim.kr

Selim 세림티에스지(주)

인재제일, ICT 프로젝트를 제일 잘 하는 기업

(주)우리아이티는 2003년 회사설립이래 최고의 인적자원과 탁월한 기술력을
바탕으로 정보시스템 및 정보통신망 구축과 유지관리, 소프트웨어개발 등
부산/울산/경남지역의 IT통합 서비스를 제공하는 회사입니다.



WIT (주)우리아이티
Leading company in IT convergence

고객 중심, 책임 있는 유지관리, 안정적인 기술 지원으로
꾸준하게 신뢰를 쌓아온 IT 전문 기업입니다.

Partners

CISCO ORACLE IBM HYOSUNG 효성인포메이션시스템 DELL
NUTANIX AXGATE TMAXSOFT SUSE vmware

부산시 동구 조방로 14 범일동, 동일타워 415호
TEL 051-637-2386 I FAX 050-5964-5555

홈페이지 http://www.wooriit.kr
문의처 management@wooriit.kr

NEOBRIX

Slogan

AUTONOMY, CONFIDENCE AND INTERACTION

Management Philosophy

행복한 구성원이 행복한 기업을 만든다. 행복한 기업이 행복한 세상을 만든다.

회사명
네오브릭스 주식회사

대표이사
현종일

임직원수 (2025년 기준)
22명(연구개발/기술인력 19명)

사업영역
소프트웨어 개발, 공급 및 유지보수
시스템 통합 자문 및 구축 관련 서비스 용역











NEOBRIX

공공기관 SI 전문기업

열정과 도전정신

긍정적 사고와 혁신

인재 중심 조직 문화

고객 중심의 가치 창조

고객성공이
최우선인
"正道 경영"

최고의 인재 육성

미래지향적인
사업역량 확보

검찰청 및 공공기관 출판 및 디자인 전문 **성민기업**

공공 SI 전문기업 **나날에스엠아이(주)**



법무부	• 대법원 • 검찰청 • 대한변호사협회
환경부	• 한국환경공단 • 환경산업기술원
산업통상자원부	• 특허청 • 특허정보원 • 국세청 • 한국산업기술평가관리원 • 한국산업기술진흥원
행정안전부	• 국가정보자원관리원 • 기획재정부 • 한국재정정보원 • 한국정보화진흥원 • 교통안전공단 • 한국스마트카드

성민기업
도 서 출 판

초심을 잃지 않는 기업!

나날에스엠아이(주)
NaNal System Maintenance & Integration

| 02-571-5700 | 02-3462-4100 |

Akamai

API 가시성 확보로 조직을 위협으로부터 방어

아카마이 API 시큐리티와 함께라면 API, 더 이상 위험하지 않습니다

기업 내부/외부의 모든 API 트래픽을 추적해 보안 관점에서
API 포스처 관리와 증가하는 API 오남용에 대한 가시성을 보안팀에 제공하고
위협에 대응하며 개발팀과 효율적으로 협업할 수 있는 환경을 제공합니다.

전체 API 자산 확인

지속적인 검색과 모니터링을
통해 섀도우, 좀비, 로그 API를
포함한 모든 API를 찾아
인벤토리 작성

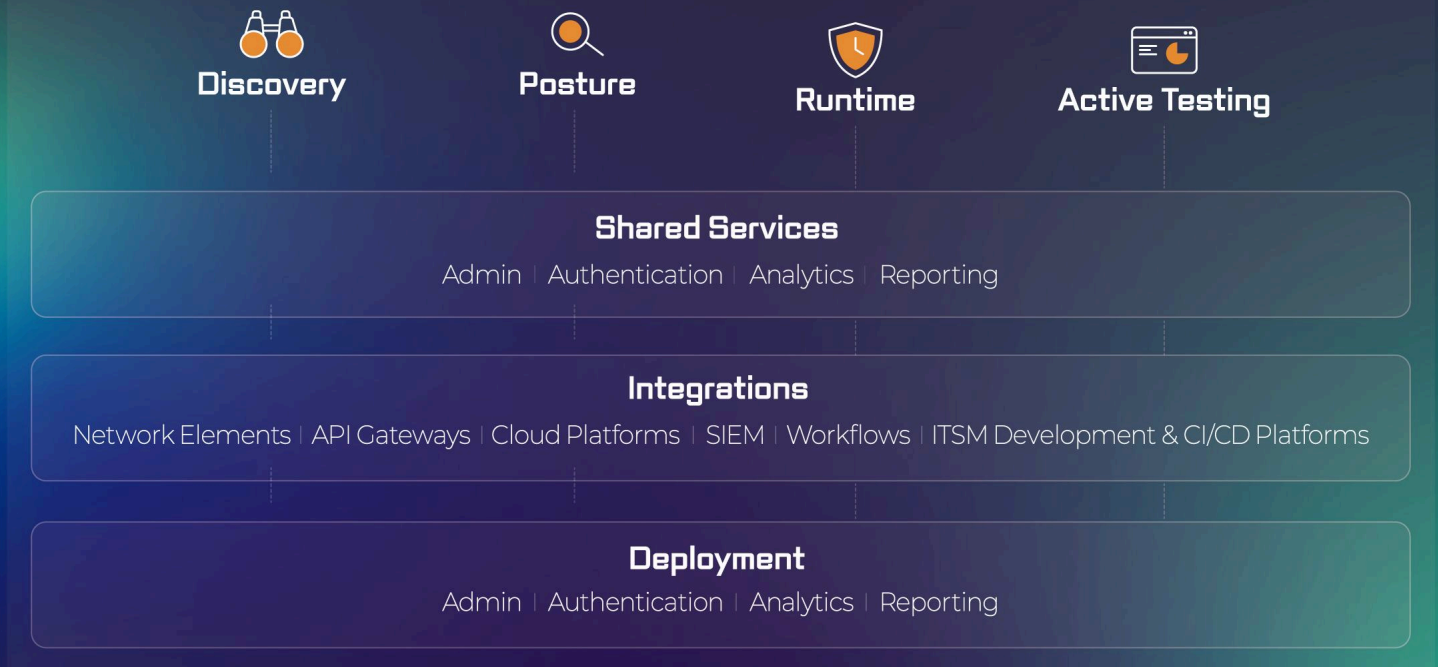
취약한 API 식별

공격자가 표적으로 삼는
API 취약점과
잘못된 구성에 대한 감사

* OWASP API 상위 10개 항목 모두 포함

API 남용 및 공격 방지

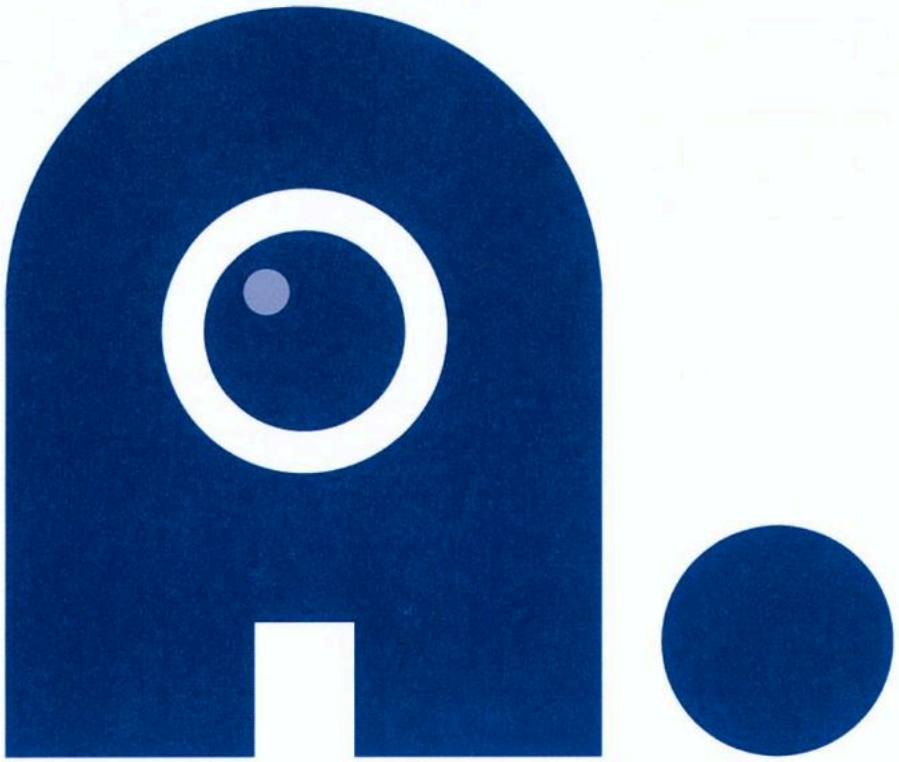
데이터 유출, 의심스러운
행동, 악의적인 봇,
API 공격 등의 위험 식별



GiT
굿모닝아이텍(주)

굿모닝아이텍(주)는 Akamai의 파트너 최고 등급인 **엘리트(Elite) 파트너사**입니다.

10545 경기도 고양시 덕양구 향동로 217 DMC플렉스데시앙 B동 7F
T. 02. 3775. 2347 | E. goodmkt@goodmit.co.kr | www.goodmit.co.kr



안녕하세요, 에이닷입니다

SK텔레콤의 새로운 AI 서비스
에이닷을 소개합니다

아직은 작고 서툴 수도 있지만
에이닷에겐 소중한 꿈이 있습니다

고민하지 않아도
좋아하는 콘텐츠를 바로 보고
말하지 않아도
필요한 순간에 힘이 되어 주는 노래를 듣고
애쓰지 않아도
원하는 것을 자유롭게 할 수 있는
새로운 세상을 여는 꿈

에이닷과 시간을 보내는 사이
우리는 이 새로운 꿈에
매일 가까워지고 있습니다

최고의 AI를 넘어 Art로 성장하는
에이닷의 꿈에 함께해 주세요

일상의 디지털 메이트,
에이닷

SK telecom

인천퍼스트치과

FIRST DENTAL CLINIC

평생구강 관리 관심있다면
'인천퍼스트치과'

본점	보미점
032) 429-2275	032) 439-2755
연수점	**구월점**
032) 214-2875	032) 465-2755



ITCEN ENTEC

Inspire with Technology

세상을 변화시키는 기술

우리는 고객에 대한 이해를 넘어서,
변화의 흐름을 읽고, 멈추지 않는 도전을 통해
지식과 기술로 세상을 혁신합니다.

깊은 신뢰감으로 늘 고객과 함께하고,
기술, 사람, 기업간 새로운 연결 속에서
새로운 경험과 가치를 발굴하며
아이티센은 지속가능한 변화를 실현합니다.

더 나은 내일을 향한 기술로
고객의 삶에 필요한 가치를 창조하는 것.
아이티센이 생각하는
우리의 역할이자 우리의 존재 이유입니다.

