

인공지능의 형성: 윤리, 권력, 그리고 책임

SHAPING AI: ETHICS, POWER, RESPONSIBILITY

인공지능의 형성: 윤리, 권력, 그리고 책임

보고서

날짜:

2025. 11. 28 - 11. 29

장소:

주한독일문화원, 용산구 소월로 132, 서울특별시, 대한민국

주최자:

안드레아 벨라피(Andrea Bélafi),
레슬리 클라테(Leslie Klatte), 주한독일문화원

작성자: 최소영(Soyoung Choi), Daily prompt

교역자 및 번역자:

추영롱(Young-Rong Choo), 번역가 | 문화 기획자 | 활동가

영어 번역: 에릭 로젠크란츠(Eric Rosencrantz)

독일어 번역: STAR Korea AG

비주얼: 오주영, 예술가/연구자

편집, 디자인 및 워크숍 사진:

레슬리 클라테(Leslie Klatte), 주한독일문화원

Day 1

모더레이터: 이창희, KAIST 산업디자인학과 교수

퍼실리테이터:

장여경(Yeo-Kyung Chang), 정보인권연구소 상임이사
권유빈(Yubeen Kwon), 서울대학교 박사과정
오주영(Jooyoung Oh), 연구 기반으로 시각, 기술, 사회를 탐구하는 예술가
가야 나다라잔(Gaya Nadarajan), SKKU 문헌정보학과 부교수
염인화(Inhwa Yeom), 미디어 아티스트, 연구자, BiOVE 창립자

Day 2

모더레이터: 박주용, KAIST 문화기술대학원 교수

퍼실리테이터:

I.T. 신(I.T. Shin), 연세대학교 정치외교학과 겸임교수, OGQ Corp.CEO
천현득(Hyundeuk Cheon), 서울대학교 과학학과 교수
전유진(Youjin Jeon), 다학제적 예술가, Woman Open Tech Lab 대표
한울(Oul Han), 서강대학교 글로벌한국학과 조교수
김인숙(In Suk Kim), ADeKo 과학기술이사

목차

1. 배경	2
2. 서론	3
3. 워크숍 요약	4
4. 주요 발견 및 권고안	18
5. 퍼실리테이터 소개	19
6. 참고 자료	22

1. 배경

2024년 12월, 윤석열 전 대통령이 사실상 쿠데타 시도라고 볼 수 있는 계엄령을 선포했을 때, 대한민국은 중대한 정치적 위기를 경험했다고 UCLA 러스킨 공공정책대학원(2025)은 밝혔다. 비록 윤 전 대통령의 계획은 수포로 돌아갔고 그 역시 파면되었지만, 이 사건은 한때의 일제 식민지, 전쟁의 폐허에서 기적을 만들어 낸 국가, 급속한 경제 성장과 함께 권위주의 정부를 거쳐 융성한 민주주의를 발전해 온 대한민국이라는 한 국가 내에 잔류하는 취약성을 드러내었다.¹

오늘날, 대한민국은 민주적 책임성, 공공 서비스 제공 및 국가적 역량의 측면에서 높은 위상을 차지하고 있다. 하지만 개발 및 민주주의에 관한 중차대한 어려움은 여전히 상존하고 있다. 여러 국제 비교 분석에 따르면, 한국은 정치적으로 가장 양극화된 국가 중 하나에 속한다. 국가 주도의 개발은 비록 큰 성공을 거두었지만, 동시에 민주적 통합을 저해하는 구조적 유산을 남겼다. 특히 취약한 정당 체제와 제한된 시민 사회를 그 특징으로 꼽을 수 있다. 이러한 잔재는 뿌리 깊은 사회적 분열을 심화한다. 그리고 친위 쿠데타 시도로 드러난 정치적 균열뿐만 아니라, 외교 정책과 경제 우선순위, 그리고 젠더 문제에 관한 첨예한 이견도 지속되고 있다. 실제로 한 연구에² 의하면, 한국의 정치 및 사회적 환경은 세대 갈등, 젠더 갈등, 지역 갈등 등으로 깊이 분열해 있음을 확인할 수 있다.

UCLA 러스킨 공공정책대학원(2025)은 한국의 불안정한 지정학적 위치 때문에 특히 외교 정책이 매우 논쟁적인 주제라고 강조한다. 미·중 갈등과 결합한 북한의 안보 위협은 전략에 관한 첨예한 이견들을 양산한다. 보수층은 미국과의 강한 동맹 및 평양에 대한 강경책에 찬성하는 반면, 진보 진영은 관계 개선을 지지한다. 한국의 최대 교역국인 동시에 북한의 주요 우방국이기도 한 중국, 그리고 한국의 기존 안보 동맹 사이에서 외교 노선을 정하는 일은 국내 정치에서 가장 양극화된 주제가 되었다.

마찬가지로, 젠더에 기반한 양극화 역시 분열의 주된 요인이 되었다. 구조적이고 가부장적인 요인들이 이러한 의견 차이를 유발하는데, 여기에는 한국이 OECD 국가 중 젠더별 임금 격차가 가장 심한 국가 중 하나인 점, 남성의 병역 의무, 급속한 고령화 사회, 그리고 인구 급감을 야기할 정도로 세계 최저 수준인 출생률이 포함된다. 이러한 조건들은 경제적 지속 가능성, 사회적 응집력, 미래 노동력 안정성에 높은 압박을 가하고 있다.

아울러, 부와 기회는 여전히 균등하지 않게 분배되어 있다. 대한민국은 소득 불평등 지수에서 OECD 국가 중 상위 50%에 속하며, 청년 세대가 도시로 이주하며 더 심화된 도시-농촌 간의 격차를 직면하고 있다. 제한된 사회적 이동성과 널리 확산한 비관주의적 태도는 특히 젊은 세대에 영향을 미치는데, 이들은 선진국 중 가장 긴 노동시간을 가진 한국의 고강도 노동 문화 속에서 점점 더 희망을 잃어간다.

그렇지만 이 같은 어려움 속에서도, 대한민국은 국내총생산의 상당 부분을 연구개발에 할당하고 있으며, 이는 OECD 국가 중 2위에 해당한다. 2022년 R&D 투자는 경제 총생산의 5%를 초과하였는데, 이 수치는 10년 이상 유지되고 있다. 이처럼 지속적인 투자는 인공지능을 포함한 첨단 기술 분야에서 한국을 글로벌 리더의 자리로 이끌었다.³

본 보고서는 급속한 기술 발전이 한국에 어떤 새로운 책임과 위험을 초래하는지 이해하는 데 핵심적인 맥락을 제공한다. 인공지능이 경제·사회·정치 체제 속에 깊숙이 자리 잡기 시작하면서, 윤리와 거버넌스, 책임에 관한 문제는 더욱 시급해지고 있다.

혁신은 발전을 추동하지만, 통제되지 않은 혁신은 불평등을 심화하고, 편향을 강화하며, 사생활을 침해할 수도 있다. 이러한 기반을 바탕으로, 본 보고서는 대한민국의 인공지능 개발이 윤리적 원칙에 부합하도록 보장하는 데에 있어 직면한 주요 과제를 검토하며, 혁신과 사회적 신뢰 및 공정성 간의 균형을 맞추는 틀을 구축하기 위한 전략을 모색한다.

1 Edward L. Knudsen, Helmut K. Anheier and Joseph C. Saraceno, [Progress and Polarization in South Korea: Remarkable Economic Ascent Masks Lingering Fractures](https://www.luskin.ucla.edu), UCLA Luskin School of Public Affairs, Los Angeles, 2025. (luskin.ucla.edu)

2 Asia Pacific Foundation of Canada. 2024. Power Struggles and Polarization Plunge South Korea into Crisis. Available at: <https://www.asiapacific.ca/publication/power-struggles-and-polarization-plunge-south-korea-crisis>

3 Edward L. Knudsen et al., op. cit.

2. 서론

인공지능(AI)은 한국의 산업, 교육, 미디어, 일상생활 전반에서 활용 범위를 빠르게 확장하며, 개인의 생활 방식과 조직의 업무 수행, 사회적 의사결정 절차에까지 영향을 미치고 있다. 이러한 변화에 대응하기 위해 한국은 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」을 제정하여(법률 제20676호) 2026년 1월 22일 시행을 예고하고 있다. 또한 과학기술정보통신부는 해당 법이 2024년 12월 26일 국회 본회의를 통과했음을 공식적으로 알린 바 있다.

그러나 법제의 정비가 곧바로 사회적 신뢰의 확보로 이어지는 것은 아니다. AI의 개발과 운영은 종종 소수의 대규모 기업 및 제한된 정책 결정 구조에 집중되는 경향을 보이며, 그 결과로 혜택과 위험이 사회 내부에서 비대칭적으로 배분될 가능성이 제기된다. 더불어 AI 시스템은 기존의 데이터와 제도, 문화적 규범 위에서 작동하기 때문에 성별, 사회적 배경, 문화적 차이와 관련된 편향을 재생산하거나 강화할 위험 또한 내포한다. 따라서 한국적 맥락에서 ‘책임 있는 AI’의 조건을 논의하는 일은 기술 진흥의 논리를 넘어, 권력과 책임의 배치, 공정성과 참여의 원칙을 함께 점검하는 과제가 된다.

이러한 문제의식 속에서 주한독일문화원은 ‘인공지능의 형성: 윤리, 권력, 그리고 책임(Shaping AI: Ethics, Power, Responsibility)’ 워크숍을 개최하였다. 본 워크숍은 한국적 맥락에서 AI의 윤리적, 사회적, 기술적, 정치적 차원을 조명하고, 교육, 미디어, 제도, 예술을 가로지르는 논의를 통해 한국 AI 생태계의 미래와 윤리적 프레임워크 구축에 관한 쟁점을 구체화하는 데 목적을 둔다. 또한 이 워크숍은 주한독일문화원(Goethe-Institut Korea)과 독일연방시민교육청(Bundeszentrale für politische Bildung)이 함께 추진하는 국제 AI 프로젝트의 일환으로, 각 지역의 경험과 쟁점을 교차 비교하며 공공적 논의를 확장하는 장을 지향한다.

본 워크숍은 다음과 같은 질문들을 중심으로 논의를 전개한다.

- AI 시스템에서 공정성, 투명성, 책임성을 어떻게 보장할 것인가. 동시에 공동체가 자체 데이터 세트를 생성하고 관리할 수 있도록 지원하는 조건은 무엇인가.
- AI를 통제하는 주체는 누구인가. 글로벌 빅테크와 로컬 혁신가의 역할은 한국의 디지털 주권 논의에 어떤 함의를 갖는가.
- AI는 어떤 방식으로 공공과 함께 만들어지고, 공공을 위해 작동하는 공동의 문화 자원이 될 수 있는가.
- 민주적 가치와 문화적 규범을 반영하는 책임 있는 AI 개발을 위해 정책결정자, 개발자, 시민은 각각 어떤 역할을 수행해야 하는가.
- 비전문가가 유해한 AI 요소를 이해하고 검증하며 때로는 거부할 수 있도록, 어떤 교육적·법적 프레임워크가 필요한가.

프로그램은 국내 AI 연구자, 윤리 및 사회과학자, 기술 전문가, 예술가가 참여하는 발표, 토론, 참여형 세션으로 구성되었으며, 한국어로 진행되었다. 본 보고서는 이러한 논의 구조를 바탕으로, 한국 사회에서 책임 있는 AI를 구축하기 위한 핵심 쟁점과 향후 과제를 정리한다.

3. 워크숍 요약 // 첫째 날 요약



2025년 11월 28일 진행된 워크숍
키오프 발표 현장. 장여경(왼쪽),
사회자 이창희(오른쪽).

인권과 한국의 인공지능법

장여경 (정보인권연구소 상임이사)

장여경은 인공지능을 규제 완화와 산업 진흥의 관점이 아니라, UN의 기업과 인권 이행지침(UNGPs)을 바탕으로 한 ‘인권기반 접근’으로 다루어야 한다고 강조했다. 콜센터 노동자 해고, 교사와 학생들의 딥페이크 피해 사례 등을 제시하며, AI 환경에서는 피해의 경계가 모호하고 개별 피해자가 이를 입증하기 어려운 구조가 형성되고 있음을 지적했다. 예를 들어, 자율주행차가 흰색 트럭을 ‘구름’으로 오인한 사고처럼, AI 시스템의 오작동은 운전자·센서·알고리즘·학습 데이터 중 어느 부분에 책임이 있는지 가리기 어려워 법적 책임을 묻는 데 한계가 있다고 했다.

또한, 한국 인공지능법이 고위험·고영향 AI를 규제 대상으로 삼고 있음에도, 생체인식·감정인식·감시 시스템과 같이 인권 침해 가능성이 큰 기술을 명시적으로 금지하지 않으며, 국방·안보 목적의 AI는 아예 규제 대상에서 제외되어 있다고 설명했다. 사고가 발생할 경우에도 모델 개발사, B2B 사용자, 최종 이용자 사이에 책임이 분산돼 피해자가 실질적으로 구제받기 어려운 점을 지적했다.

이를 보완하기 위해 인권영향평가의 법제화와 자료 공개 의무를 제안했다. AI가 시장에 나오기 전에 어떤 집단에 어떤 영향을 미치는지 평가하고, 사고 시 피해자가 기업·기관에 설명을 요구하고 이의를 제기할 수 있는 절차를 법에 담아야 한다는 것이다. 그는 AI로 인해 영향받는 모든 사람을 권리 주체로 규정하고, 기업과 정부가 투명성과 책무성을 법적 의무로 이행해야 한다고 강조했다.

인공지능 윤리 탐구: ‘이루다’ 사례와 인간중심 인공지능

권유빈 (서울대 박사과정)

권유빈은 한국의 챗봇 ‘이루다’ 사례를 통해 ‘작동하는 윤리(ethics-in-action)’ 개념을 제시했다. 이루다 1.0은 연애 컨설팅 서비스의 카카오톡 데이터를 활용하면서 개인정보보호와 혐오 발화 문제로 큰 사회적 논란을 겪고 서비스가 중단된 바 있다. 이후 기술·시장·규제기관·사용자·윤리 지침을 순환하는 다섯 개의 루프를 통해 다시 사회적으로 수용 가능한 기술로 복귀하였다. 권유빈은 이 과정을 분석하면서, 추상적 원칙이 아닌 실제 상황에서 작동하고 전이 가능한 ‘중범위 개입’으로서의 윤리가 중요하다고 주장했다.

이어 최근 각국과 산업에서 핵심 키워드로 부상한 ‘인간중심 인공지능(Human-centered AI)’의 역사와 철학을 소개하며, 인간 개입과 통제, 투명성과 설명 가능성을 중심에 두는 접근이 한편으로는 탈인간주의·생태 윤리·데이터 노동에 대한 성찰을 충분히 담지 못하고 있음을 비판했다. 그는 이에 대한 대안으로 ‘관계중심 인공지능’을 제안하며, 인간을 고정된 개체로 전제하기보다 다양한 존재들이 맺는 관계와 그 관계가 유지되는 생태계의 지속 가능성을 설계의 중심에 두어야 한다고 주장했다. 이때 ‘페미니스트 AI 프레임워크’는 특정 의제라기보다, 관계적 존재론을 기반으로 착취적 관계를 해체하고 상호적 관계를 구축하는 이론적 토대로 언급되었다. 예컨대 모델을 개발할 때 잠재적 사용자의 편의만이 아니라 언어 공동체나 소수 집단 등 관계망 전체에 미칠 영향을 함께 검토해야 하며, 관계 생태계를 훼손하는 기술이라면 개발 중단까지도 논의할 수 있어야 한다는 점이 강조되었다.

권유빈(왼쪽)이 ‘생활 윤리(lived ethics)’와 한국 챗봇 이루다 사례에 대한 분석을 발표하고 있다.



포스트 감성 시대의 현실: 편향, 몰입, 포용적 인공지능

오주영 (연구 기반으로 시각, 기술, 사회를 탐구하는 예술가)

오주영은 사진·미디어 아티스트이자 연구자로서, AI를 ‘모델’보다 ‘데이터 세트와 아카이브’의 문제로 다뤘다. 그는 ‘이미지 유사도 네트워크’를 통해 시각적 패턴을 분류하고, 그 과정에서 ‘무엇이 대표로 보존되고 무엇이 누락되는가’를 드러내는 작업을 이어왔다. 또한 저작권 만료 문학 텍스트를 활용한 챗봇 프로젝트를 소개하며, 기록과 재현이 기술의 문제가 되기 이전에 권력과 선택의 문제였음을 짚었다.

작가는 서울시립 사진미술관으로부터 개관 기념 작품 제작을 의뢰받아, 1920년대부터 1980년대까지 한국 사진계가 중요하게 간주해온 이미지들을 생성형 AI로 ‘복원’해보는 실험을 시작했고, 동시에 이미지 유사도 네트워크를 구성해 ‘유사할수록 가까이, 전체 데이터셋에 비해 특이할수록 멀리’ 배치되는 시각적 지형을 살폈다. 분석 과정 중 네트워크에서 다른 이미지들과 유난히 떨어져 있는 ‘특이한 데이터’가 눈에 띄었는데, 그것은 구분창 작가의 사진이었다. 작가는 사전 정보를 반영하지 않은 상태에서도 특정 작가의 특수성이 시각적 패턴으로 포착될 수 있음을 확인했으며, 예컨대 구분창 작업들에서 반복되는 ‘빨간 점의 위치’ 같은 트레이드마크가 유사도로 묶이며 드러났다고 설명했다.

그는 이어 관람객이 이미지를 선택해 입력하면, AI가 이를 해석해 비평 형태의 텍스트를 출력하는 참여형 설치 작업 <기계 감상 프로그램>을 언급했다. 겉으로는 인간의 감상이 AI에게 학습 신호를 주는 것처럼 보이지만, 실제로는 우리가 AI의 비평에 일희일비하고 그 평가를 마치 맞는 말처럼 느끼며 넘어가 버리는 순간들이 쌓인다는 것이다. 그러다 보면 인간 자신의 감상 능력을 빼앗기는, ‘환영에 가까운 상태’가 만들어지게 된다. >



오주영이 AI를 ‘데이터 세트와 아카이브’로 바라보는 자신의 예술 연구를 발표하며, 이미지 유사도 네트워크와 한국 사진의 생성적 재구성 작업을 함께 소개하고 있다.

작가는 <기계 감상 프로그램>에 함께 전시된 ‘사진 복원 측정 차트’도 이어서 다루었다. 필름 사진을 복원하고 디지털 아카이브로 옮기는 과정에는 이미지의 상태와 차이를 점검하고 비교하는 차트가 존재하며, 눈으로 보는 사진과 화면에서 빛의 값으로 보이는 사진이 얼마나 비슷하고도 다른지 측정하는 데 사용된다. 다만, 작가가 만든 차트는 실제 아카이빙 작업에서 쓰이는 차트 형식만 가져왔을 뿐, 실제로는 존재하지 않으며 “유머러스하게 불가능에 가까운 방식으로 왜곡시킨 작업”이라고 설명했다. 예를 들어, 작가가 제작한 ‘AI USAF1951 패턴’은 실재하지 않는 초해상도 그룹을 만든다는 그럴듯한 기능을 지녔지만 실제로는 존재하지 않는 차트라는 것이다.

그의 발표에서 중요한 축은 <문학소녀 챗봇 1974>였다. 이는 돌아가신 할머니가 사실은 작가였음을 뒤늦게 알게 된 개인적 경험에서 출발해, 여성 작가의 텍스트가 어떤 방식으로 남거나 사라지는지를 데이터 세트 구축의 문제로 연결한다. 과거 신문 기사 이미지는 ‘책 읽는 여성’이 위험하다는 사회적 분위기를 보여주는 자료로 제시됐다. 광주에서 상위권 여고생 세 명이 내장산으로 떠나 동반 자살했다는 보도 및 그 원인을 “앙드레 지드 같은 문학책을 읽으며 세상을 비판했기 때문”으로 돌리는 해석은 당시의 생활 세계를 드러낸다. 그는 이런 맥락 속에서 할머니가 글을 쓰고 남기기 어려웠을 조건들을 구체적으로 떠올리게 된다고 말했다.

<문학소녀 챗봇 1974> 프로젝트를 설명하며, 무엇보다 ‘원문을 어떻게 다룰 것인가’를 먼저 고민했다고 말했다. 작업은 원본을 새로 쓰거나 변형하는 방향이 아니라, ‘철저하게 원본을 변형하지 않고’ 텍스트를 재배열하는 방식에 가깝게 설계되었다. 작가는 이런 접근을 통해 기술이 개입해야 할 지점과 멈춰야 할 지점을 분명히 구분해야 한다고 주장했다.

마지막으로 그는 여성의 서사가 텍스트에서만 빠지는 게 아니라, 이미지에서도 비슷한 방식으로 왜곡되고 누락된다고 짚었다. 1970년대 한국, 박정희 정권 시기에 발행된 여학생 잡지의 ‘표지’만을 학습한 시는 ‘여학생’이라기보다 그 시대가 보여주고 싶었던 ‘소비되는 여성의 이미지’를 반복해서 보여주었다. 그리고 ‘여학생·문학소녀’라는 주제를 연구해 온 박숙자 교수에게 이러한 이미지를 보여주자, “이거는 여학생이 아니죠”라는 반응이 돌아왔다고 한다. 그래서 오주영은 여학생지에서 사진 콘테스트를 열어 여학생들이 자신을 담아 찍었던 사진들을 함께 제시하였고, 이런 자료를 더 많이 보여줄 수 있는 곳이 있으면 좋겠다며 발표를 마무리했다.

이미지 유사도 네트워크와 아카이브 데이터를 통해 AI에 접근하는 오주영의 작업 이미지. 1920~1980년대 한국을 대표하는 사진들을 비교·‘복원’한 서울시립미술관(SeMA) 위촉 프로젝트 작업.





가야 나다라잔이 생성형AI가 한국 대학 교육을 어떻게 변화시켰는지에 대한 자신의 분석을 발표하고 있다.

생성형 인공지능과 윤리: 학생들은 어떻게 생각할까?

가야 나다라잔 (Gaya Nadarajan, SKKU 문헌정보학과 부교수)

가야 나다라잔은 코로나 이후 온라인 수업과 플립 러닝이 정착된 상황에서 생성형 AI가 교육 현장에 미치는 영향을 분석했다. 그는 ChatGPT와 같은 도구가 등장하면서 학생 과제의 완성도가 급격히 높아졌고, 학업 성실성과 비판적 사고가 위협받고 있다는 점을 지적하였다. 나다라잔 교수 또한 개인적으로 강의 자료를 급히 준비할 때 AI에 의존하게 되며 스스로 사고하는 능력이 감소한다고 느꼈다.

이러한 변화에 대응하기 위해 그는 강의-실습-과제 제출 방식 대신, 수업 후 학생들에게 추가 질문을 통해 이해도를 평가하는 방법을 도입하였다. AI 활용 여부에 대한 명확한 제도적 규정이 부재한 상황에서, 한국과 캐나다 대학생들을 대상으로 한 비교 연구를 수행하였다. 연구에서 캐나다 학생들은 동일한 사례에서도 AI 사용을 윤리적·규정적 위반으로 인식하는 경향이 강하게 나타났으며, 한국 학생들은 이를 더욱 관대하게 평가하는 것으로 나타났다. 캐나다 학생들은 과제 결과를 ‘개인의 독창성과 책임’으로 증명해야 한다는 규범이 강해, 같은 상황에서도 AI 사용을 규정 위반이나 부정행위로 해석하는 경향이 강했다. 반면 한국 학생들은 지식과 학습을 공동의 과정으로 보는 경향이 있어, AI를 ‘협력적 학습 보조’에 가까운 일종의 도움으로 받아들이며 윤리적 부담을 상대적으로 적게 느꼈다.

다만 두 집단 모두 AI가 생성한 코드 비중이 커질수록 더 비윤리적이라고 판단하는 경향을 보였으며, 나다라잔은 이를 교육 설계의 핵심 신호로 보았다. 따라서 그는 일률적 금지나 전면 허용보다, 문화적 인식 차이를 고려한 명확한 가이드라인과 함께 ‘어디까지가 도움이고 어디부터가 책임 회피인지’를 학생 스스로 판단할 수 있도록 비판적 사고와 책임 있는 의사 결정을 훈련하는 교육이 필요하다고 강조했다.



염인화가 AI 기술의 수명 주기, 자원 분배, 그리고 생성형 AI에서의 아시아 여성 재현 문제를 논의하고 있다.

인공지능을 활용한 미디어 생성: 질문과 과제

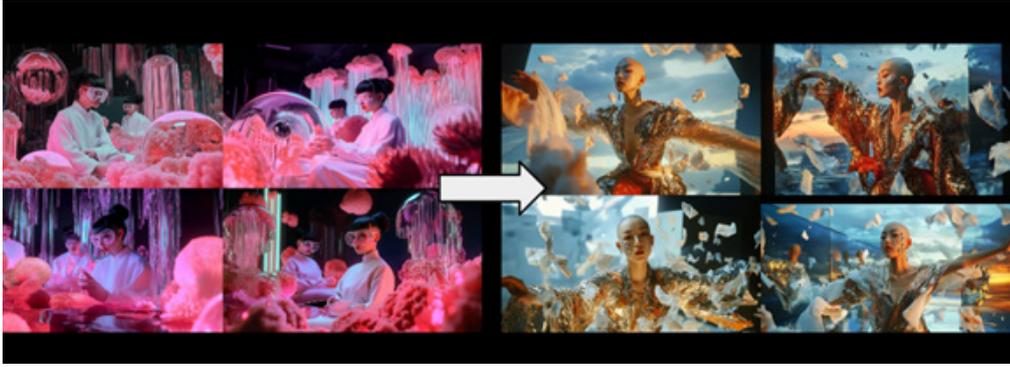
염인화 (미디어 아티스트, 연구자, BiOVE 창립자)

염인화는 확장현실과 공연예술에 기반하고 생성형 AI를 활용해온 경험을 바탕으로, AI 기술의 수명주기와 자원 분배 문제를 짚었다. 그는 기술 수용 곡선이 더 이상 ‘누가 먼저 들어왔는가’로 구분되지 않을 정도로 빠르게 재편되는 현실에서, AI는 제작 시간과 인력을 줄여주지만 그만큼 자본과 권력이 집중되는 ‘자원의 전쟁’을 심화시키고 있다고 분석했다.

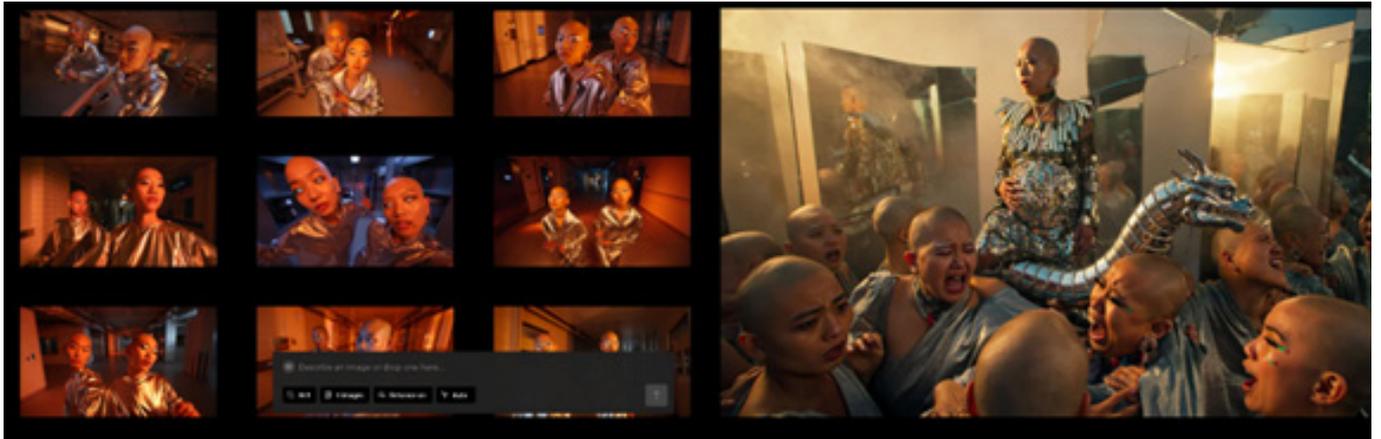
이 관점은 ‘재현’의 문제로도 이어졌다. 작가는 아시아 여성 재생산 문제를 다루는 작업 과정에서, 생성형 AI로 아시아 여성을 만들면 그 결과가 K-pop 아이돌 같은 이미지, 이상화되고 왜곡되고 성적대상화되는 여성 이미지가 대량으로 나오는 현상을 직접 겪었다고 말했다. 그 스타일이 모델의 ‘기본값’처럼 작동하면서 아시아 여성이 ‘소비 가능한 젊은 몸’으로 단일화되기 쉽다는 문제를 지적했다. >



Midjourney AI에 ‘한국 여성’을 입력해 생성된 이미지 Midjourney AI에 ‘아시아 여성’을 입력해 생성된 이미지



Midjourney AI로 생성된 <War Dance> (2025) 테스트 이미지 (염인화 작). 마른 체형, 하얀 피부, 메이크업, 시선과 포즈 등 미화되거나 '서구화된' 또는 '성적 대상화된' 여성 이미지의 특징을 보인다.



Reve AI로 생성된 <War Dance> (2025) 테스트 이미지 (염인화 작). '현실적인' 이미지를 생성하는 것으로 알려진 AI 모델이다.

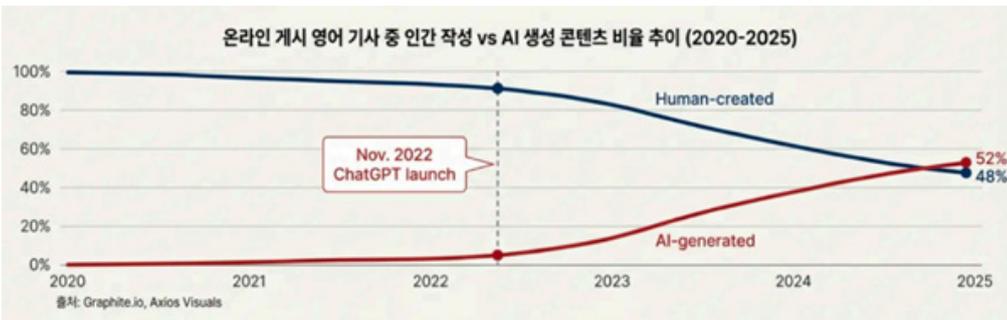
그래서 그는 대머리, 피어싱처럼 통상적 여성 이미지에서 벗어난 요소를 프롬프트에 넣어 왜곡된 결과를 우회하는 '추가 통제 노동'이 필요했다고 설명했다.

그는 AI를 쓰지 않는 선택지가 거의 사라진 상황에서, '누가, 무엇을 위해, 어떤 조건에서' AI를 쓰는지 핵심이라고 지적했다. 나아가 작업 공정 단축으로 생긴 자원을 다시 페미니즘·기후위기·신경다양성 등 동시대 담론에 투자하는 방식으로 AI를 자원 재분배의 도구로 사용할 수 있다고 제안했다.

—

1일 차 전반을 통해 참가자들은 인공지능을 인권과 법제의 문제, 데이터와 편향의 문제, 예술과 교육의 문제로 동시에 바라보며, 기술의 편의와 효율성에 가려져 있던 구조적 불평등과 보이지 않는 노동, 생태적 비용을 전면에 드러냈다. 이 논의는 2일 차에 이어지는 '거버넌스, 책임, 문화 산업과 AI 생태계' 논의를 위한 중요한 기반이 되었다.

3. 워크숍 요약 // 둘째 날 요약



I.T. 신이 한국 AI 주권 문제와 저품질 AI 생성 콘텐츠가 데이터 생태계에 미치는 영향 확대에 대해 설명하고 있다.

Trend in the ratio of text written by humans to AI-generated content in online English-language articles (2020–2025). Graphite, “More Articles Are Now Created by AI Than Humans” (2024), quoted from Axios: “Exclusive: AI writing hasn’t overwhelmed the web yet” (2025)

인공지능이 제시하는 과제와 기회

I.T. 신 (연세대학교 정치외교학과 겸임교수, OGQ Corp.CEO)

I.T. 신은 한국의 ‘AI 주권’을 둘러싼 과제와 기회를 다루었다. 그는 우선 국가가 공공데이터를 개방하기 위해 구축한 플랫폼이 존재하지만, 실제로 올라온 데이터는 수량도 제한적이고 내용도 지엽적이어서 현장에서 바로 쓰기 어려운 수준에 머문다는 점을 짚었다. 그 사이 인터넷, 소셜미디어, 블로그, 이미지와 영상까지 AI에 의해 생성되거나 부분적으로 의존해 만들어지는 비중이 빠르게 커지고 있으며, 출판 시장에서도 AI로 대량 생산된 책이 유통 구조로 밀려들어 기존의 교정·검수 시스템이 품질을 걸러내지 못하는 사례가 나타난다고 진단했다. 그는 이런 ‘품질이 낮은 대량 생성물’의 확산을 ‘AI slop(찌꺼기)’이라는 말로 요약하며, 데이터 생태계가 오염될수록 다음 학습의 입력값이 더 왜곡되는 악순환이 빨라진다고 평가했다.

이어 AI 윤리 부재가 차별의 자동화를 초래하는 사례로, 정치·사회적 중립의 불가능성과 편향된 데이터에 기반한 모델이 ‘일반적인 사람’, ‘다정함’, ‘연쇄살인범’ 등을 어떻게 묘사하는지 실험 결과를 통해 보여주었다. 그 결과 ‘전형적인 사람’은 특정 인종·성별·직업으로 수렴하고, 다정함은 특정 외모·표정에, 폭력성은 백인 남성 이미지에 과도하게 연결되는 등, 서구 중심 데이터와 안전 규칙이 만들어낸 기본값이 드러났다.

마무리에서 그는 편향 문제는 ‘해결’이 아닌 ‘완화’를 목표로 뒤야 한다며, 한국 사회의 가치와 맥락을 반영한 윤리 데이터 세트를 신속히 구축할 필요가 있다고 강조했다. 또한 기업과 공공기관이 모델을 도입·운영할 때 프라이버시와 책임성, 사회적 위험을 점검할 평가 세트와 스크리닝 도구가 필요하다고 제안하며, 이를 위한 웹 기반 평가 프로토타입을 시연했다.

인공지능, 책임의 공백, 그리고 AI 블랙박스 문제

천현득 (서울대학교 과학학과 교수)

천현득은 ‘인공지능 윤리, 책임의 간격, 블랙박스 문제’를 주제로, AI 윤리 가이드라인들이 공통적으로 강조하는 투명성, 인간 중심, 책임성을 검토했다. 그는 공정성 시비에 민감한 사회일수록, 오히려 눈에 보이는 객관적 절차와 기술 지표에 과도하게 의존하려는 경향이 생긴다고 지적했다. 그러나 어떤 데이터를 선택하고 어떤 방식으로 학습할지 결정하는 순간부터 가치판단은 불가피하며, 기술이 ‘더 객관적’이라는 인식 자체가 위험할 수 있다고 경고했다. 그는 자율 무기와 같이 전쟁 범죄가 발생했을 때 누구도 책임을 지지 않는 상황을 예로 들며, 인간의 통제와 인식 조건이 무너질 때 생기는 ‘책임/응보의 공백’을 설명했다.

또한 젠나 버렐의 세 가지 불투명성(의도적, 문해력, 알고리즘 불투명성)을 통해 블랙박스 문제를 분석하며, 특히 공공 정책에 사용되는 AI 시스템은 시민이 수용할 수 있는 이유로 정당화되어야 하므로 설명 가능성과 공공성을 강하게 요구해야 한다고 강조했다.

천 교수가 투명성, 책임성의 공백, 기술적 객관성 신화 등 AI 윤리의 핵심 과제를 설명하고 있다.

그는 기술이 자연재해처럼 ‘어쩔 수 없이 다가오는 것’이 아니라, 사회 구성원들의 다양한 가치가 절충된 결과물이라는 점을 상기시키며, ‘미래 비전을 초국적 빅테크 기업에 외주화하지 말 것’을 당부했다.





전유진이 한국에서의 AI 이해를 위한 핵심 질문들을 제시하고 있다.

AI 과잉 소비로 우리가 잃는 것

전유진 (다학제적 예술가, Woman Open Tech Lab 대표)

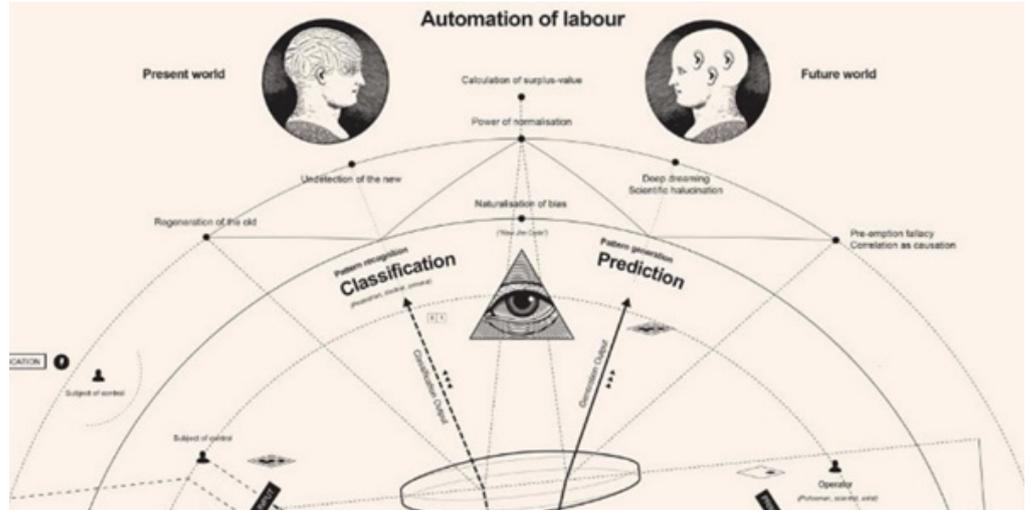
전유진은 강연 초반에, AI를 논할 때 먼저 짚고 넘어가야 할 배경적 문제의식을 정리하며 출발했다. AI는 서비스이자 상품이고, 우리가 접하는 AI는 이미 생활 속으로 대중화된 소비 대상이기 때문에, ‘무엇을 얻는가’만큼 ‘무엇을 잃는가’를 소비의 관점에서 점검해야 한다는 문제의식이었다. 그는 이를 위해 먼저 다섯 가지 배경을 짚었다. 기술은 종립적이지 않으며 편향 위에서 작동한다는 점, “왜 인공 ‘지능’인가”라는 명명 자체가 상징과 기대를 만든다는 점, AI가 특정한 한 기술(LLM 기반 챗봇)로 환원될 수 없다는 점, AI는 오래된 기술이면서 계속 변하는 기술이라 블랙박스화와 소외를 낳기 쉽다는 점, 그리고 기술 발전 속도, 상품화 속도, 담론 형성 속도 사이의 간극이 너무 커, 충분한 논의 없이 AI 사용이 당연한 선택이 되어버린다는 점이다.

편향의 문제는 젠더화된 분류 사례로 구체화되었다. 전유진은 2015년 Max Dovey의 컴퓨터 비전 기반 인식 실험 작업을 예로 들며, 남성의 몸조차도 나체일 때 ‘여성’으로 분류되는 사례를 소개했다. 즉, 사례를 통해 ‘나체는 모두 여성’이라는 기술의 편향성을 짚었다. 이러한 오류는 단순한 오인식이 아니라, 데이터가 ‘몸’을 어떤 관습으로 분류해왔는지와 연결된다. 학습 데이터와 라벨링 과정이 성별을 이분법으로 고정하고, 노출된 신체를 ‘여성’이라는 범주와 강하게 결합해 온 시각문화의 규칙을 그대로 흡수한 것이다.



How to be more or less human by Max Dovey (2015)

The "Nooskoop Manifesto": Counter-narratives to the AI hype (13th Gwangju Biennale)



이 문제의식을 압축적으로 보여주는 장치로 13회 광주비엔날레에 소개된 블라단 올러와 마테오 파스퀴넬리의 ‘누스코프 선언’을 인용했다. ‘누스코프 선언’은 AI를 ‘지능적 기계’라는 이념적 지위에서 사회를 읽고 탐색하는 ‘지식도구’로 다시 배치하려는 시도라고 설명했다.

또한 기계학습을 어떤 신비한 능력으로 떠받들기보다, 인간의 손이 닿기 어려운 방대한 데이터 공간에서 특징과 패턴, 상관관계를 포착하도록 돕는 ‘지식확대경’으로 이해하는 편이 타당하다고 덧붙였다. 이런 맥락에서 ‘지능’이라는 말이 기술을 신비화하는 방식이 드러나고, AI는 초월적 존재라기보다 사회를 측정하고 분류하는 장치로 다시 읽히게 된다. 전유진은 “AI의 신비적 요소를 제거할 필요성은 기업에서도 알고 있다”고 덧붙이며, 현재의 AI를 ‘인지(cognition)’라기보다 고도화된 ‘인식(perception)’에 가까운 기술로 보는 안 르쿤(메타 인공지능 연구소 초대 소장)의 관점도 함께 소개해, ‘지능’이라는 단어가 만들어내는 과장된 기대를 걷어낼 필요가 있다고 강조했다.

이어서 전유진은 “왜 이것을 인공 ‘지능’이라고 부르는가”라는 언어·상징 차원의 문제를 제기했다. 그는 AGI, ‘일반’ 인공지능, 메타버스 등 화려한 이름이 실질적 기능과 과장된 기대를 뒤섞어, 기술을 하나의 거대한 블랙박스로 만들어 버리는 효과를 낳는다고 분석했다. 특히 OpenAI가 AGI를 ‘일반적으로 인간보다 더 똑똑한’ 것으로 정의할 때 ‘general(일반적)’이라는 말은 정상/비정상 같은 규범의 경계를 떠올리게 하기 때문에, 그는 보편성을 가장한 언어가 배제의 기준으로 작동할 수 있음을 경계했다. 아울러 ‘생각한다’, ‘이해한다’, ‘거짓말한다’ 같은 인간 중심의 동사가 AI에 붙는 순간, 그 자체가 이미 경계 신호라고 강조했다.

기술을 도구로 다루기보다 ‘행위자’처럼 오해하게 만들고, 그 오해가 과잉 소비와 의존을 촉진할 수 있다는 판단이다. 또한 생성형 AI가 데이터를 기반으로 만들어 내는 만들기(generating)와 인간이 읽고 사유하고 자신만의 관점을 만들어나가는 주체적인 앎의 과정(becoming)은 질적으로 다르다는 점을 강조하며, 빠른 ‘생성’에 의존할수록 진정성과 자기성찰을 놓치게 될 위험을 지적했다.

왜 인공 '지능'인가? 적합한 이름인가? 'AGI' 등 AI와 관련한 여러 어휘와 상징



OpenAI는 그들의 임무를 설명하면서 AGI에 '일반적으로 인간보다 더 똑똑한(generally smarter than humans)'이라는 정의를 덧붙이지 않았는가. 그들은 AGI의 general이라는 수식어를 이미 이런 식으로—인간과 대립하는 위계적 방식으로—쓰고 있다. 위에서 말한 '범용'이라는 뜻으로 국한하지 않는다. 그리고 떠오른 질문은 '일반에 미치지 못하는 AI를 쓰는 이들은 누구일까?', '일반 인공지능을 쓰지 못하면 일반에 속할 수 없는 건가?'처럼 '일반'이라는 수식어는 자연스럽게 사용자를 향하는 쪽으로 확장되었다. 그렇게 보면 general은 어떤 존재를 범주화하는 힘이 있고, 그 집합에 포함될 수 없다는 두려움을 자극하는 언어다. AI의 흐름에 함께 하지 않으면 다수에 포함될 수 없다는 두려움을 자극하는 것이 AGI 속 general의 용도라고 말한다면 비약일까?

「AI를 위한 새로운 이름 짓기」_전유진 (2023)
<http://semacoral.org/features/youjinjeon-renaming-ai-against-labeling-technology>

이 지점에서 그의 논의는 “그러면 우리는 이 기술을 어떤 말로 불러야 하는가”라는 실천 문제로 확장된다. 전유진은 명명 행위가 피할 수 없는 인간의 습성임을 인정하면서도, 자본과 권력이 결합한 기술 언어가 지식 체계를 흔들고 위계를 재생산하는 지금, 예술이 할 수 있는 일은 허황되고 위험한 어휘를 골라내 ‘정직한 이름’을 다시 붙이는 일이라고 제안한다.

그는 마지막으로 기술 발전 속도와 상품화 속도, 그리고 이에 대한 사회적 담론이 형성되는 속도 사이의 간극을 강조했다. 기술의 변화가 속고보다 빠르게 기본값이 되어버리는 상황에서, 소비자는 수동적 사용자(‘호구’)가 아니라, 서비스를 중단 또는 불매하고 대안 도구를 선택함으로써 적극적으로 개입할 수 있는 주체라고 주장했다. 더 늦기 전에 우리가 되돌리고 다시 점검해야 할 것은 성능 경쟁 이전에, ‘언어가 만든 기대’가 ‘소비의 구조’로 굳어지는 방식이라고 결론지었다.

K-콘텐츠와 산업 AI 생태계 연결 가능성

한울 (서강대학교 글로벌한국학과 조교수), 김인숙 (ADeKo 과학기술이사)

한울은 K-콘텐츠와 산업 AI 생태계를 연결해 분석했다. 글로벌 동시 공개와 빠른 생산·유통을 전제로 하는 한국 콘텐츠 산업에서, 번역·자막·팬 커뮤니케이션·마케팅 문구 제작 등은 이미 AI 없이 유지되기 어려운 인프라가 되었음을 보여주었다. 그는 LLM 기반 번역과 카피라이팅이 점차 정서·문화 표현 방식의 ‘기본값’을 형성하며, 어떤 감정 표현이 적절한지, 어느 정도의 거리감이 읽기 쉬운지 등 새로운 정서 규범을 제안하고 반복한다고 설명했다. 이 과정에서 문화적 차이가 평평하게 늘리는 ‘문화적 평탄화(cultural flattening)’의 위험이 존재한다고 지적했다. 그는 한국이 풍부한 사례와 실험을 제공하는 역할을 한다면, 독일과 같은 국가들은 기준과 규범을 함께 설계해 나갈 수 있기 때문에, 초기부터 국제 협력과 네트워크 구축이 중요하다고 강조했다.

한울(왼쪽)은 AI가 K-콘텐츠의 급속한 글로벌 확산을 어떻게 뒷받침하는지 분석하며, LLM 기반 번역과 카피라이팅이 변화하는 정서·문화적 규범에 어떠한 영향을 미치는지 보여주었다.

AI와 한국의 도시·공공 데이터 생태계의 상호작용을 조명한 김인숙(오른쪽).

김인숙은 도시 데이터와 공공 데이터 사례를 통해, 문제는 데이터의 부재가 아니라 신뢰의 부재와 의사결정 권한의 집중이라고 분석했다. 즉, 데이터가 순환하지 못하는 문제의 근본적 원인은 데이터를 보유한 기관과 그 데이터를 실제로 활용하고자 하는 주체 사이에 신뢰가 형성되지 않은 점에 있으며, 그 신뢰가 생기지 않는 배경에는 데이터에 참여하는 사람들이 자신의 데이터를 통제하고 활용할 결정권을 충분히 갖지 못하는 현실이 놓여 있다고 보았다. 그는 산업 AI 비즈니스에서 공급자와 소비자가 직접 만나기 어렵기 때문에 중간자 역할이 필요하다고 했고, 어떤 영역이든 공급자 중심으로만 설계할 것이 아니라, 소비자·이용자·납세자까지 포함해 이해관계자를 재구성해야 한다고 강조했다. 또한 기술자·경제학자·인문학자(커뮤니케이터)가 함께 뛰어야 한다고 제안했다.



마무리로 한국과 독일의 협업을 이야기하며, 한국은 디지털 기반이 촘촘하고 속도와 시각화가 강점인 반면, 독일은 ‘브레인과 아키텍처’ 같은 설계 역량이 강하다고 비교했다.

그래서 독일의 설계력과 한국의 실행력이 결합될 때 결과가 빨라지고 성과가 뚜렷해지며, 이 조합을 바탕으로 글로벌 확장까지 노려볼 수 있다고 정리했다.

—

마지막 종합 토론에서는 ‘좋은 AI’의 조건과 시기, 그리고 인간의 주체성을 둘러싼 질의응답이 진행되었다. 연사들은 자율성, 책임성, 투명성, 공공성을 핵심 가치로 제시하며, AI를 인간의 역량을 강화할 수 있는 도구로서 위치 지어야 한다고 강조했다. 그러나 알고리즘 추천과 초개인화 마케팅, 감시 자본주의의 구조 속에서 인간의 취향과 주체성이 범주화·정량화되고, 기본소득과 같은 제도가 도입될 경우 ‘생각하기를 포기한 다수’와 ‘소수의 개성 추구자’로 사회가 양극화될 수 있다는 우려도 제기되었다. 또한 햇볕과의 대화가 나르시시즘을 강화하고 마찰 없는 관계를 제공함으로써, 타자성과 갈등을 통해 성장하는 인간의 경험을 약화할 수 있다는 지적도 나왔다.

마지막으로, 한 참가자는 도구의 역사와 함께 to have(소유), to do(능력·직업), to be(존재·의미)라는 세 가지 축을 제시하며, AI 인프라가 결국 향하는 곳은 어디인지 질문했다. 이에 대해 연사들은 산업사회가 주로 to have와 to do를 지원하는 도구를 만들어 왔다면, 이제는 to be, ‘우리는 누구이며, 어떻게 살 것인가’라는 질문에 답하는 예술·인문학·철학이 더 중요해질 수 있다고 응답했다. 둘째 날 논의는 결국 AI를 둘러싼 규범과 생태계를 설계하는 과정에서 기술·산업의 논리만이 아니라 인간의 의미와 존엄을 함께 사유해야 한다는 점으로 수렴되었다.

11월 29일 박주용의 사회로 진행된 마지막 패널 토론에서는 자율성, 책임성, 투명성, 공공 참여의 중요성을 강조하며 ‘좋은 AI’와 인간의 주권을 위한 조건에 대해 논의했다.



4. 주요 발견 및 권고안

본 워크숍은 한국에서 윤리적 AI가 기술 성능이나 추상적 가치 선언의 문제가 아니라, 권리와 책임, 데이터와 재현, 교육과 사용 규범, 그리고 사회적 신뢰의 구조와 긴밀히 연결되어 있음을 확인하는 자리였다. 또한 UCLA 러스킨 공공정책대학원 (2025)에서 지적하듯 한국은 세계적으로 높은 교육 수준을 유지하고 있으며, 이러한 교육 기반은 혁신을 견인하는 인적 자본일 뿐 아니라 시민 참여와 민주적 회복탄력성을 가능하게 하는 조건으로 작동할 수 있다. 연사들의 논의는 이러한 사회적 역량이 윤리적 AI 논의의 중요한 배경임을 전제로 삼고 있었다.

여러 연사의 발언을 통해 공통적으로 확인한 첫 번째 쟁점은, 윤리적 AI가 선언적 원칙이나 자율적 가이드라인에 머물 경우 실제 피해를 예방하거나 구제하는 데 한계가 크다는 인식이었다. 장여경은 고영향 AI 영역에서 예외와 회색지대가 확대될 수록 기본권 침해 위험이 커지며, 피해가 발생했을 때 책임과 구제의 경로가 불분명해질 수 있음을 지적했다. 천현득 역시 블랙박스 시스템과 에이전트형 AI의 확산이 책임 주체를 흐리게 만들고, 설명 가능성이 결여된 자동화가 공공 영역의 정당성을 약화할 수 있음을 강조했다. 이러한 논의는 윤리적 AI가 가치 선언의 문제가 아니라, 책임과 설명, 구제 절차가 실제로 작동하는 구조의 문제임을 드러낸다.

데이터와 재현의 문제 또한 기술 외곽이 아닌 핵심 쟁점으로 반복적으로 언급되었다. 오주영은 데이터 세트와 아카이브가 무엇을 기록하고 무엇을 누락하는지가 재현의 윤리를 결정하며, AI 복원과 생성이 ‘그럴듯함’을 강화할수록 오류와 누락이 더욱 자연스럽게 고착될 위험이 있음을 지적했다. 염인화는 생성형 AI의 확산이 효율과 생산성의 언어로 설명되는 동안, 자원과 권력이 누구에게 집중되고 비용이 누구에게 전가되는지에 대한 질문이 충분히 제기되지 않는 현실을 비판했다. 이 두 논의는 윤리적 AI가 편향을 기술적으로 제거하는 문제를 넘어, 누락과 불균형이 어떻게 구조화되는지를 함께 성찰해야 함을 시사한다.

교육과 일상적 사용의 맥락에서도 윤리적 AI의 쟁점은 분명하게 드러났다. 가야 나다라잔은 교육 현장에서 생성형 AI가 학습을 지원하는 동시에 성실성과 비판적 사고를 약화할 수 있으며, 문화적 배경에 따라 AI 사용을 인식하는 기준이 다르게 형성된다는 점을 강조했다. 전유진은 기술을 의인화하는 언어가 이미 사회적으로 경계해야 할 신호로 작동하고 있으며, 이러한 언어 사용이 과잉 기대와 의존을 강화할 수 있음을 지적했다. 한울의 논의는 K-콘텐츠 산업에서 번역과 자막, 팬 커뮤니티가 이미 AI 기반 인프라로 자리 잡았고, 이 과정에서 감정 표현과 문화적 차이가 평탄화될 위험이 존재함을 보여주었다. 이러한 발언들은 윤리적 AI가 기술 사용의 허용 여부가 아니라, 어떤 사용 규범과 문화적 기준이 형성되고 있는가의 문제로 이해되어야 함을 드러낸다.

윤리가 실제로 작동하기 위한 조건으로서 운영과 평가의 문제 역시 중요한 공통 인식으로 나타났다. 권유빈은 ‘이루다’ 사례를 통해 윤리가 기술 외부의 선언으로만 존재할 경우 쉽게 무력화될 수 있기 때문에, 서비스의 설계와 운영 전반에서 반복적으로 점검되고 갱신되어야 함을 강조했다. I.T. 신은 편향을 완전히 제거하기보다 완화하는 것이 현실적 목표이며, 이를 위해 윤리 데이터 세트와 평가 세트, 스크리닝 도구와 같은 검증 장치가 필요하다고 설명했다. 이러한 논의는 윤리적 AI가 의도의 문제가 아니라, 측정과 점검이 가능한 운영 구조의 문제임을 시사한다.

이러한 연사들의 논의를 종합할 때, 본 워크숍은 한국에서 윤리적 AI를 둘러싼 쟁점이 개별 기술이나 단일 제도의 문제가 아니라, 다층적 네트워크와 공론장을 통해 다루어져야 함을 보여준다. 연사들은 각기 다른 영역에서 권리 보호, 데이터와 재현, 교육과 사용 규범, 평가와 운영의 중요성을 강조했으며, 이러한 쟁점들이 분절된 채로는 해결되기 어렵다는 점을 공통적으로 드러냈다. 이 맥락에서 문화기관과 같은 중간 조직의 역할, 즉 정책, 산업, 시민사회 사이의 논의를 연결하고 서로 다른 맥락과 언어를 번역하며 논의를 기록하고 축적하는 기능의 필요성이 자연스럽게 부각되었다. >

마지막으로, 배경에서 언급된 정치·사회적 분열은 윤리적 AI 논의를 어렵게 만드는 조건으로 작용한다. 분열된 환경에서는 윤리 기준이 쉽게 진영의 언어로 소비되고, 기술 정책에 대한 신뢰가 약화할 수 있다. 동시에 연사들의 논의는 투명성, 설명, 책임과 같은 절차적 요소가 기술에 대한 ‘결과의 합의’가 아니라 ‘과정의 정당성’을 공유하는 기반이 될 수 있음을 시사한다. 한국의 높은 교육 수준과 시민 참여 역량은 이러한 절차를 이해하고 검증하는 사회적 조건으로 작동할 수 있으며, 본 워크숍은 윤리적 AI 논의가 어떠한 쟁점과 조건 위에서 전개되고 있는지를 구체적으로 보여주었다.

5. 퍼실리테이터 소개



이창희 Changhee Lee

이창희 교수는 KAIST 산업디자인학과에 재직 중이며, KAIST 우주연구소와 AI연구원에서 활동하고 있다. 그의 연구는 실험적 인터페이스와 인터랙티브 시스템을 통해 감각 경험을 탐구한다. 영국왕립예술학교(RCA)에서 박사학위를 받았으며, RCA와 임페리얼 칼리지 런던이 공동 운영하는 혁신디자인공학 프로그램에서 강의했다. UK 디자인위원회가 선정한 Ones to Watch로 주목받았고, Prix Ars Electronica Digital Humanity 부문(2025)에서 명예언급을 받았으며, 영국문화원 동문상 최종 후보에 올랐다.



장여경 Yeo-Kyung Chang

장여경은 한국 비영리 단체 진보넷(디지털 권리 연구소)의 사무총장이다. 그는 인권 관점에서 디지털 기술과 사회의 관계를 연구하며, 한국 정책 담론에 디지털 권리 개념을 도입하는데 핵심적인 역할을 해왔다. 이전에는 개인정보보호위원회 비상임 위원을 역임했으며, 현재 한국국가인권위원회 디지털 권리 전문가 위원회 위원을 맡고 있다.



권유빈 Yubeen Kwon

권유빈은 서울대학교 과학기술학 박사과정에 재학 중이며, 동 대학에서 교육학 학사 및 석사 학위를 취득했다. 연구 관심사는 전문성, 인간-컴퓨터 상호작용, 인공지능 윤리, 시민 교육 등이다. 최근 『Social Studies of Science』에 인공지능 챗봇 이루다의 부상과 몰락을 분석한 논문을 게재했다. 박사 연구에서는 인공지능 시대에 과학적 전문성이 어떻게 재구성되고 있는지 탐구하며, 특히 인공지능-생물학 협업 연구 그룹을 중심으로 연구를 진행 중이다.



오주영 Jooyoung Oh

오주영은 시각성, 기술, 사회의 교차점을 탐구하는 연구 기반 예술가이다. 그의 작업은 인공지능, 시각 패턴, 인식 시스템을 활용해 인간과 기계의 지각이 어떻게 상호작용하는지 살펴보고 예술과 기술의 새로운 가능성을 열어간다. 최근 그는 기후 위기, 이동성, 환경 자원, 돌봄과 같은 주제에 집중하면서 누락되거나 주변화된 데이터의 윤리적 문제를 비판적으로 다루고 있다. 그의 작업을 통해 간과된 이야기와 보이지 않는 데이터가 드러나며, 기술을 통해 형성되는 우리 세계에 대한 새로운 시각을 제시한다.



가야 나다라잔 Gaya Nadarajan

가야 나다라잔(Gaya Nadarajan)은 성균관대학교(SKKU) 부교수로, 유럽과 아시아에서 25년 이상 연구와 산업을 연결해 온 경험을 보유하고 있다. 그의 전문 분야는 영상 분석, 신약 개발을 비롯해 최근에는 교육 기술 및 연구에 이르기까지 다양하다. 가야는 한국, 캐나다, 영국 출신 학생 및 교수진 간의 지적 교류와 문화 간 교류를 촉진하는 협력적 온라인 교육 이니셔티브에 기여해 왔으며 그의 연구는 데이터 과학, 교육, 글로벌 협업의 교차점에서 혁신을 주도하고 있다.



염인화 Inhwa Yeom

염인화는 미디어 아티스트이자 연구원이며 BiOVE의 창립자이다. XR 및 AI 기술과 공연 예술의 교차점에서 그는 생물 식민주의, 기후 위기, 신경다양성을 다루며 다양한 (비)인간적 경험을 포용한다. 염인화는 LG아트센터·LG전자가 수여하는 미디어 아티스트 상 수상자이며, 현대자동차 그룹이 주최하는 제6회 VH AWARD 최종 후보에 올랐다. 그의 작품은 국립현대미술관(MMCA), 아르스 일렉트로니카 페스티벌, ACC, SIGGRAPH Asia 등에서 전시되었다.



박주용 Juyong Park

박주용은 이론물리학 박사이며 문화, 예술, 창의성을 연구하는 복합계학자이다. 복합연결망으로부터 시작해 지금은 문명의 두 기둥이며 서로를 포용하는 ‘문화’와 ‘과학’을 진보케 하는 인간 창의성의 ‘연결성’을 연구하고 있다. 또한 창의성에 대한 통찰을 통해 기계가 진정한 지능을 가질 수 있는지 집중적으로 연구해 왔다. 서울대학교에서 물리학사, 미국 미시간대학교-앤아버에서 물리학 박사를 받은 뒤 현재는 카이스트 문화기술대학원 교수 및 한국고등과학원(키아스) 방문교수로 재직 중이다. 과학-문화 융합 베스트셀러 에세이집 “미래는 생성되지 않는다: 포스트 AI시대, 문화물리학자의 창의성 특강”(동아시아, 2024)의 저자이기도 하다.



I.T. Shin IT Shin

I.T. Shin은 OGQ Corp.의 CEO이자 연세대학교 정치학 및 국제관계학 겸임교수로 재직 중이다. KAIST에서 기술경영학 석사 학위를 취득했으며, 연세대학교에서 정치학 및 외교학을 전공했다. I.T. Shin은 창작자와 팬을 연결하는 IP 콘텐츠 플랫폼 OGQ, 환자와 전담 의사를 연결하는 맞춤형 의료 서비스 Docfriends, 정치인에 대한 일상적 평가를 통해 3선 제한제 도입을 목표로 하는 정치 혁신 플랫폼 POSDAQ 등 여러 혁신 플랫폼의 창립자이다.



천현득 Hyundeuk Cheon

천현득은 과학철학자로 과학의 발전을 성찰하고, 이를 바탕으로 철학을 확장하는 데 관심이 있다. 최근에는 인공지능의 철학적·윤리적 쟁점, 특히 감정, 이해, 설명 가능성, 투명성 등을 연구하고 있다. 서울대 물리학과를 졸업하고 동 대학원에서 과학철학으로 석사와 박사학위를 받았으며, 이화여대 인문과학원을 거쳐 2018년 서울대 철학과에 부임, 2022년 신설 과학학과로 옮겼다. 현재 서울대 AI연구원 인공지능 ELSI센터장과 한국과학철학회 부회장을 맡고 있으며, 『토머스 쿤, 미완의 혁명』, 『과학이란 무엇인가』(공저), 『인공지능의 존재론』(공저) 등 다수의 저서를 집필했다.

전유진 Youjin Jeon



전유진은 사운드 아트, 퍼포먼스, 조각, 뉴미디어 기술을 아우르는 다학제적 예술가이다. 영화 음악 작곡으로 경력을 시작한 그는 현재 예술가 집단 ‘서울 익스프레스’를 통해 서사와 기술을 탐구한다. 2017년 서울에서 여성 오픈 테크 랩(Woman Open Tech Lab)을 설립해 기술에 대한 페미니스트적 관점과 비계층적 접근을 장려하고 있다. 공용 예술 프로젝트 ‘제로 메이즈 제로(Zero Makes Zero)’를 공동 기획했으며, ‘저항을 위한 기술(The Technology for Resistance)’을 연출했다. 저서로는 『기술을 비판적으로 읽기(Reading Technology Critically)』가 있으며, 『돌봄과 노동(Caring and Working)』을 공동 집필했다. 2021년부터 시작한 출판 프로젝트 ‘페뎀 테크 토크(Fem Tech Talk)’를 통해 다양한 시각에서 기술을 비판적으로 조명하는 책과 글을 출간하고 있다.

한울 Oul Han



한울은 서강대학교 글로벌한국학과 조교수로 재직 중이며, 한류 콘텐츠의 AI 기반 문화 번역을 연구한다. 베를린 자유대학교에서 박사학위를 받고 AI·언어기술 분야 박사후 연구원을 지냈다. BTS 관련 서적과 김창남·이규택의 한국 대중음악 저서를 번역하며 프리랜서 및 팀장으로 케이컬처 번역 실무 경험을 쌓았다. 최근에는 한국적 감정이 AI와 만나 어떻게 변형되는지를 설명하는 연구 개념을 개발하고 있다.

김인숙 In Suk Kim



김인숙 박사는 AI 전략과 국제 협력 분야의 선도적인 전문가로 활동 중이다. 현재 국가인공지능전략위원회 글로벌협력 분과위원, ADeKo 과학기술이사, 그리고 한다포럼 대표직을 맡고 있다. 독일 쾰른대학교에서 경제학 학사·석사·박사 학위를 취득했으며, 독일 FNF(Friedrich Naumann Foundation) 장학생으로 선발되었다. 김 박사는 글로벌 AI 생태계 구축에 적극적으로 기여하고 있으며, EU Gaia-X 지리적 확장 워킹그룹 활동, Industrial AI 글로벌 협력 모범 사례 개발, 그리고 모빌리티·스마트시티·제조·에너지·헬스 분야의 Data Space 연결을 추진하고 있다.

최소영 Soyeong Choi



AI 뉴스레터 데일리 프롬프트(Daily Prompt)를 발행하며, 기술이 사회적 약자와 주변부의 삶을 확장하는 공공 자원으로 기능할 수 있는 가능성을 탐구하고 있다. 홍익대학교에서 디자인 석사 학위를 취득하고 현재 서강대학교에서 아트앤테크놀로지 박사 과정을 밟고 있다. AI 생성 미디어와 물리적 설치 작업을 중심으로 활동하며, 2023년 광주디자인비엔날레와 2025년 개인전 『낮게 흐르는 물』에서 작품을 선보였다. 2023년 창간 이후 독창적 인사이트와 명확한 설명으로 주목받았으며, 주요 저서로는 『데일리 프롬프트 101』(2024)과 『챗GPT, 저는 이렇게 쓰고 있어요』(2025)가 있다. 현재 호서대학교 디지털프로덕트디자인학과 겸임교수로 재직하며, 기술과 예술을 잇는 다양한 프로젝트의 기획자로 활동하고 있다.

6. 참고 자료

- AI Red Lines. "Global Call for AI Red Lines". Accessed December 18, 2025. <https://red-lines.ai/>
- Balestri, Roberto. "Examining Multimodal Gender and Content Bias in ChatGPT-4o". *arXiv* (November 28, 2024). <https://doi.org/10.48550/arXiv.2411.19140>
- Business for Social Responsibility (BSR). *Fundamentals of a Human Rights-Based Approach to Generative AI*. 2025.
- Buolamwini, Joy and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research* 81: 77–91. 2018. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Chatterji, Aaron, Thomas Cunningham, David J. Deming, Zoë Hitzig, Christopher Ong, Carl (Yan) Shan and Kevin Wadman. *How People Use ChatGPT*. NBER Working Paper No. 34255. Cambridge, MA: National Bureau of Economic Research, 2025. <https://doi.org/10.3386/w34255>
- Choi, Seung Joon. "메타버그 세계관". In *제로의 책*. Seoul: 뜻과 닷, 2022.
- Council of Europe. *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*. 2024.
- Dovey, Max. "How to be more or less human". 2015. Accessed December 18, 2025. <https://www.maxdovey.com/howtobemoreorless/>
- European Union. *Artificial Intelligence Act (Regulation (EU) 2024/1689)*. 2024.
- Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy and Madhulika Srikumar. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center for Internet & Society, Harvard University, 2020.
- Future of Life Institute. "Pause Giant AI Experiments: An Open Letter". Open letter, March 22, 2023. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- Harrington, Brian, Irina Zlotnikova, Gayathri Nadarajan and Samuel Ekundayo. "Did Alice Do Wrong? Cross-Cultural Differences in Student Perceptions of Generative AI Use in University Computing Education". *ACM Transactions on Computing Education* (2025). <https://doi.org/10.1145/3776558>
- Head, Keith Robert. "Minds in Crisis: How the AI Revolution is Impacting Mental Health". *Journal of Mental Health & Clinical Psychology* 9, No. 3 (2025). <https://www.mentalhealthjournal.org/articles/minds-in-crisis-how-the-ai-revolution-is-impacting-mental-health.pdf>
- Jeon, Youjin. "AI를 위한 새로운 이름 짓기." SeMA Coral (Seoul Museum of Art Research Lab), July 20, 2023. <http://semacoral.org/features/youjinjeon-renaming-ai-against-labeling-technology>
- Jeon, Youjin. "생산성 생성하기: 아시상의 생산성 정치와 기술 구조." In *잃어가는 것들*. Seoul: National Museum of Modern and Contemporary Art, Korea (MMCA), 2025.
- Jobin, Anna, Marcello Lenca and Effy Vayena. "The Global Landscape of AI Ethics Guidelines". *Nature Machine Intelligence* 1 (2019): 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kim, Sung-woo. *인공지능은 나의 읽기-쓰기를 어떻게 바꿀까: 지금 준비해야 할 문해력의 미래*. Paju: 유유, 2024.
- Kwon, Yubeen and Sungook Hong. "Human-Centered AI, Its Possibilities and Limitations". *Journal of Science & Technology Studies* 25, No. 1 (2025): 33–63. <https://doi.org/10.22989/jsts.2025.25.1.2>
- Ministry of Science and ICT (Republic of Korea), et al. *AI Ethics Standards: Human-Centered AI (사람이 중심이 되는 인공지능(AI) 윤리기준)*. December 23, 2020. PDF.

Morrone, Megan. "Exclusive: AI writing hasn't overwhelmed the web yet". Axios, October 14, 2025. <https://www.axios.com/2025/10/14/ai-generated-writing-humans>

National Human Rights Commission of Korea (NHRCK). *Human Rights Guidelines for AI Development and Utilization*. 2022.

Office of the United Nations High Commissioner for Human Rights (OHCHR). "A/HRC/59/32: Practical Application of the Guiding Principles on Business and Human Rights to the Activities of Technology Companies, Including Activities Relating to Artificial Intelligence". Thematic report, June 16, 2025. <https://www.ohchr.org/en/documents/thematic-reports/ahrc5932-practical-application-guiding-principles-business-and-human>

Office of the United Nations High Commissioner for Human Rights (OHCHR). "B-Tech Project: Applying the UN Guiding Principles on Business and Human Rights to Digital Technologies". Accessed December 18, 2025. <https://www.ohchr.org/en/business-and-human-rights/b-tech-project>

Oh, Jooyoung. *1974 Jangchung-dong: The Bibliophile's Secret*. 2024. Paradise Art Lab.

Oh, Jooyoung. *Aura Restoration Index 2*. 2025. Photography Seoul Museum of Art.

Pasquinelli, Matteo and Vladan Joler. "The Nooscope Manifesto". 13th Gwangju Biennale (webpage). Accessed December 18, 2025. <https://13thgwangjubienale.org/pasquinelli-joler/>

Shirky, Clay. *Here Comes Everybody: The Power of Organizing Without Organizations*. New York: Penguin Press, 2008.

United Nations Office of the High Commissioner for Human Rights. *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*. New York and Geneva: United Nations, 2011.

United Nations Office for Disarmament Affairs (UNODA). "Convention on Certain Conventional Weapons: Group of Governmental Experts on Lethal Autonomous Weapons Systems (2017)". Accessed December 18, 2025. <https://meetings.unoda.org/ccw/convention-certain-conventional-weapons-group-governmental-experts-2017>

Walsh, Toby. "Open Letter to Professor Sung-Chul Shin, President of KAIST, from Some Leading AI Researchers in 30 Different Countries". Open letter (Boycott of KAIST), March 2018. <https://www.cse.unsw.edu.au/~tw/ciair/kaist.html>

Walsh, Toby and Stuart Russell, et al. "Autonomous Weapons: An Open Letter from AI & Robotics Researchers". Open letter (issued July 28, 2015, at IJCAI 2015).

Walsh, Toby, et al. "Killer Robots: World's Top AI and Robotics Companies Urge United Nations to Ban Lethal Autonomous Weapons". Open letter (embargoed August 21, 2017; released at IJCAI 2017).

Yeom, Inhwa. *Innerveauty Spa*. 2023. 3D performative apparatus environment (PC-based VR, mobile AR, video, installation). Supported by "Futures of Listening" Sound Art Lab × Residency, Asia Culture Center. Collection: Daejeon Museum of Art.

Yeom, Inhwa. *Sauna Lab Series*. 2024. Multi-channel video installation. With support from the Daejeon Museum of Art, Taiwan Contemporary Culture Lab (C-LAB) and Hualien CCIP.

Yeom, Inhwa. *Solarsonic Band*. 2024–25. 3D performative apparatus environment. With support from the Seoul Foundation for Arts and Culture (2024) and Nam June Paik Art Center and Yongin Special City (2025).

Yeom, Inhwa. *War Dance*. 2025. Single-channel video. Finalist, 6th VH AWARD. With support from the Hyundai Motor Company.

Zao-Sanders, Marc. *How People Are Really Using Generative AI Now: The 2025 Top-100 Gen AI Use Case Report*. Filtered, 2025. <https://learn.filtered.com/hubfs/The%202025%20Top-100%20Gen%20AI%20Use%20Case%20Report.pdf>

