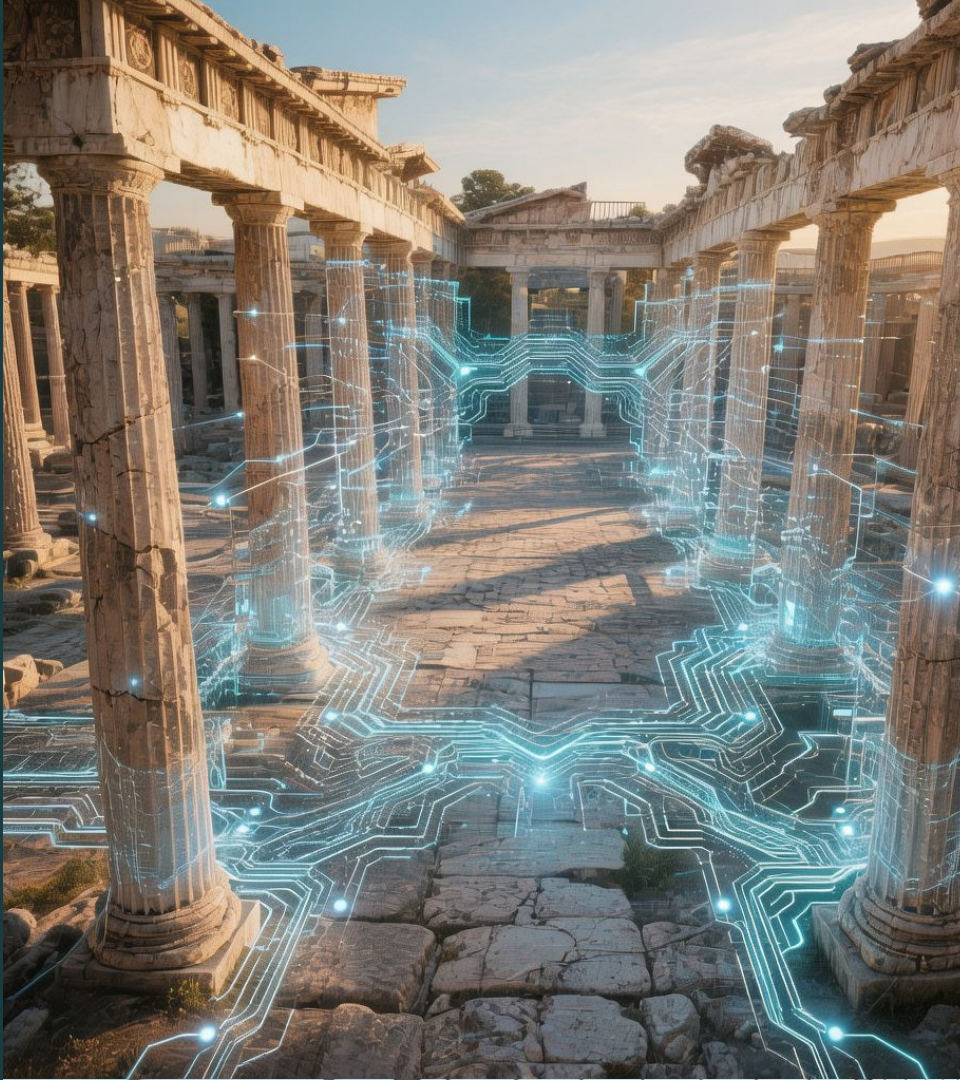


2025 KAIST The Future of AI x Humanity

# 공론장의 복원을 위하여

숙의 민주주의를 위한 인공지능, '아고라(Agora) AI' 제안

—  
전현준



AI 기술에 의해 가속화된 민주적 공론장의 훼손과 사회적 양극화

# ‘필터 버블(Filter Bubble)’ ‘반향실 효과(Echo Chamber)’

## 필터 버블 & 반향실 효과

‘사용자 참여(User Engagement)’의 극대화만을 목표로한 AI 알고리즘은

사용자의 신념을 강화하는 정보만 노출한다.

### 인지적 편향의 강화

알고리즘은 사용자의 확증 편향(Confirmation Bias)을 끊임없이 자극한다.

### 토론의 질 저하

복잡한 사회 문제는 280자의 트윗이나 1분짜리 숏폼 영상으로 소비된다.

### 신뢰의 붕괴

허위 정보와 조작된 콘텐츠가 AI를 통해 정교하게 유포되면서, 언론, 전문가, 심지어는 과학적 사실에 대한 사회 전반의 신뢰가 무너진다. ‘탈진실(Post-truth)’

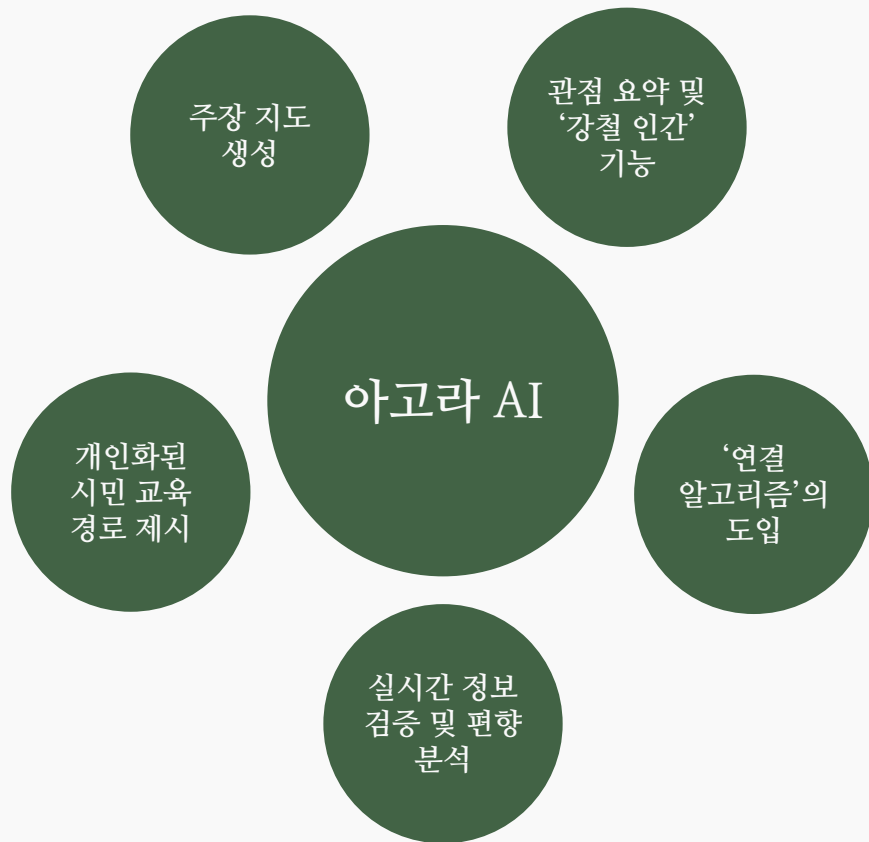


AI 기술에 의해 가속화된 민주적 공론장의 훼손과 사회적 양극화

## ‘숙의(Deliberation)’의 실종

기술이 문제를 일으켰지만,  
역설적으로 문제를 해결할 수도 있다.

# 시민들의 상호 이해와 합리적 토론을 돕는 AI 시스템



‘아고라 AI’는 정답을 제시하는 AI가 아니다.  
‘속의의 조력자’로서 기능하는 것이  
핵심 철학이다.

기대 효과

# 더 나은 사회를 향하여

## 정치

공론의 질이 향상되어 시민들이 정책 결정 과정에 더 실질적으로 참여할 수 있게 된다. 정치적 양극화가 완화될 수 있다.

## 사회/문화

혐오와 적대를 넘어, ‘다름’을 존중하고 배우려는 문화가 확산될 수 있다. 사회적 갈등을 해결하는 중요한 자산이 될 것이다.

## 경제

정치적 안정과 사회적 신뢰의 회복은 장기적으로 안정적인 경제 성장에 기여한다. 새로운 분야에서 양질의 일자리가 창출될 수 있다.



신중한 접근

## 잠재적 위험과 해결책

01

'중립성'의 편향성

오픈소스 모델 채택

독립적인 AI 윤리 위원회 감독

02

개인정보 및 사상 검증

데이터 수집 최소화 및 익명화

온디바이스(On-device) AI  
기술 활용

03

비판적 사고 퇴화

결론이 아닌 원본 자료와  
다양한 관점 제공

미디어 리터러시 공교육 연계

04

기술 접근성 격차

공공 인프라(도서관 등) 구축

디지털 활용 교육 프로그램  
지원

## 실행 경로

# 공공 인프라 구축을 위한 단계적 접근

### 글로벌 컨소시엄 구성

대학, 비영리 재단, 기업이  
참여하는 국제 컨소시엄 구성

투명한 윤리 현장 제정 및  
오픈소스 핵심 모델 개발

01

### 파일럿 테스트 및 사회적 합의 형성

대학, 언론사, 지역 커뮤니티 등  
제한된 환경에서 시범 운영

효과 및 부작용 평가, 피드백을  
통한 기술 보완 및 공감대 형성

02

### 공공 인프라로서의 단계적 확장

공공 사이트(정부, 도서관)에  
우선 적용

민간 플랫폼의 자발적 도입을  
위한 API 제공 및 인센티브  
마련

03

### 일상 속 도구로 자리매김

모든 시민의 일생 전 주기에  
활용될 수 있도록 접근성  
고도화

04

인류의 공존과 번영에 기여

## ‘신뢰의 회복’과 ‘집단 지성의 강화’

### 다시 함께 생각하는 법

서로 다른 의견을 가진 사람들이 최소한의 공통된 사실 기반 위에서 대화하고, 상대방의 입장을 이해하려는 노력을 기울일 때, 사회적 신뢰라는 자본이 축적된다.

### 인류의 집단 지성의 촉매

현재의 AI가 개인의 지능을 확장하는 ‘도구’에 머무른다면, ‘아고라 AI’는 개인과 개인이 연결되어 만들어내는 ‘관계의 지능’, 즉 집단 지성을 강화하는 ‘환경’으로서 기능한다.

### 단순히 기술적인 제안을 넘어

AI를 인간의 이성을 대체하고 통제하는 방향이 아니라, 인간의 이성을 보완하고 상호 간의 이해를 돕는 방향으로 발전시킬 때, 비로소 기술은 더 나은 미래를 위한 동반자가 될 수 있다.

## VALUE PROPOSITION

우리는 더 빠르고, 더 편하고,  
더 자극적인 것을 추구하며  
파편화된 개인으로 고립될 것인가?

## VALUE PROPOSITION

아니면 조금 더디고, 조금 불편하더라도,  
서로를 이해하고 함께 숙의하며  
더 나은 공동체를 만들어 갈 것인가?

# 공론장의 복원을 위하여

숙의 민주주의를 위한 인공지능, ‘아고라 AI’