

AUGUST 12, 2024

Final Exam
Ph.D. in Law

Collaborative Approaches to AI Governance

Exploring Co-Design and Co-Regulation Models

INYOUNG CHEONG

icheon@uw.edu

COMMITTEE

Tadayoshi Kohno (Chair)

Ryan Calo

Hugh Spitzer

Amy X. Zhang

Benjamin Mako Hill (GSR)

ACKNOWLEDGEMENT



TECH POLICY LAB
UNIVERSITY of WASHINGTON

 **FULBRIGHT**
Scholar Program

 OpenAI
Democratic Inputs to AI

TABLE OF CONTENT

- 1 Introduction
- 2 Theoretical Foundations
- 3 Case Study 1: Co-design
- 4 Case Study 2: Co-regulation
- 5 Paths Forward on AI Governance

1

Introduction

PROBLEMS

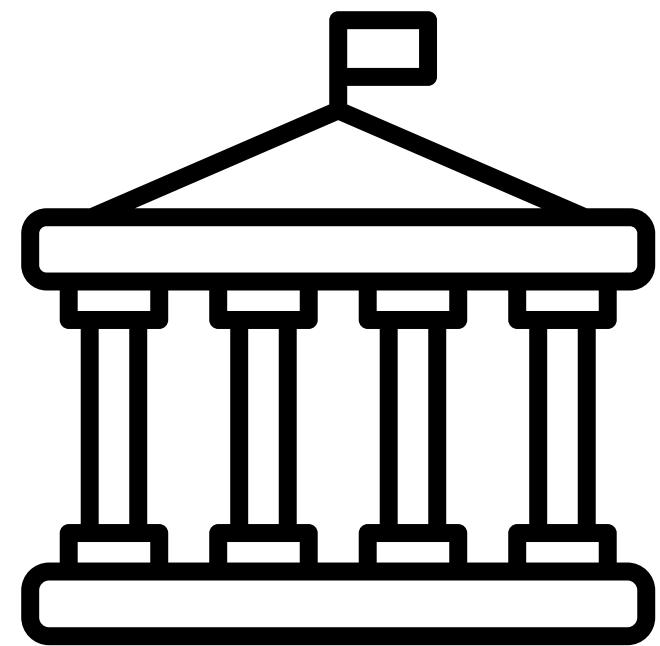
REMARKABLE CAPABILITIES OF AI TECHNOLOGIES

Human-like conversation, Video/image generation, Massive data processing, Task automation...

BROAD HARMS POSED BY AI TECHNOLOGIES

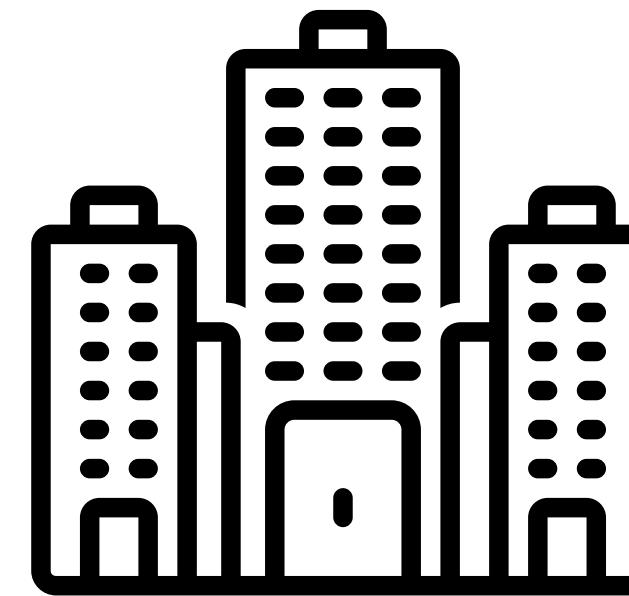
- Perpetuation of existing biases and stereotypes
- Concentration of power and authority
- Privacy erosion and increased surveillance
- Loss of autonomy
- Environmental and financial costs
- Copyright infringement and liability issues
- Job displacement
- Challenges to creativity and original human expression
- Potential for malicious use or weaponization

TRADITIONAL APPROACHES



GOVERNMENT REGULATION

VS.



MARKET FREEDOM

ALTERNATIVE APPROACHES



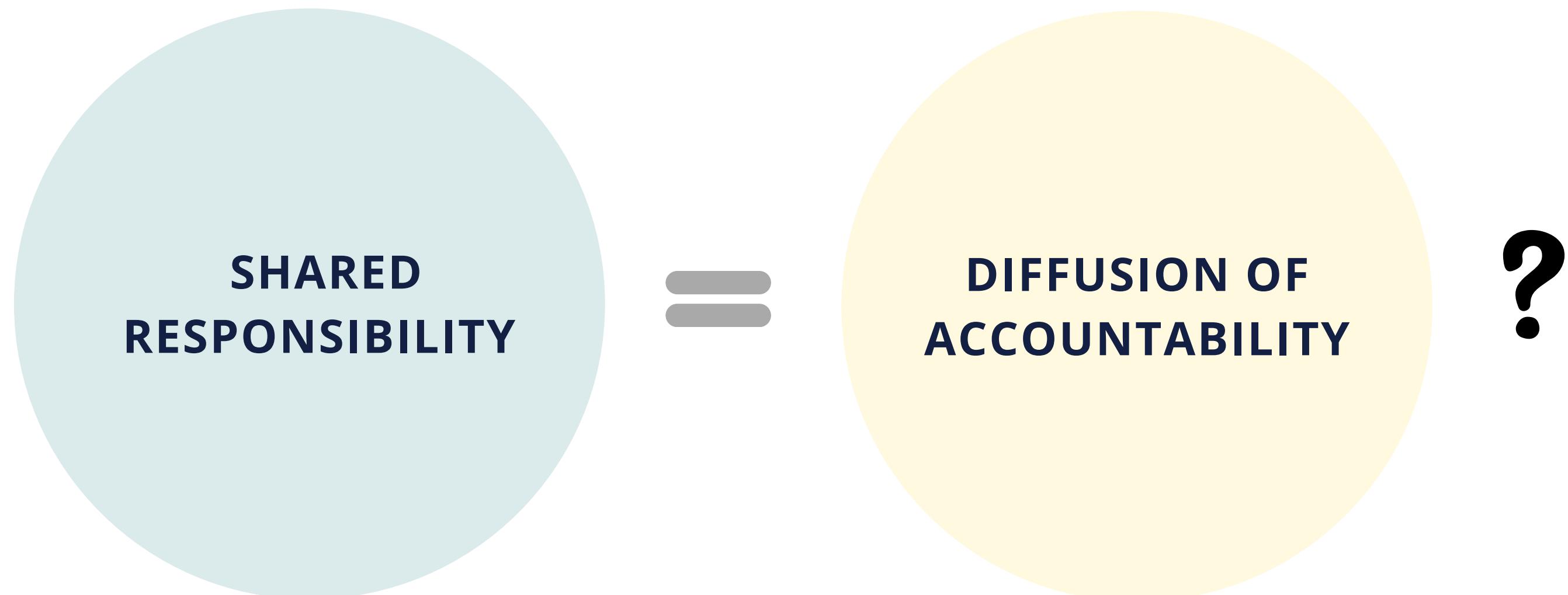
**Collabroative
rule-making & implementation**

Government actors
NGOs
Companies
Users
Communities
International agencies
Business associations ...

POTENTIAL BENEFITS

- **Increased adaptability:** More responsive to rapidly changing technologies and market conditions.
- **Innovation-friendly:** Preserves space for innovation while still providing necessary safeguards.
- **Improved information flow:** Addresses information asymmetry between regulators and industry.
- **Diverse representation:** Allows for more informed decision-making by incorporating diverse perspectives.
- **Shared responsibility:** Distributes the burden of regulation across multiple actors.
- **Trust-building:** Foster trust between industry, government, and the public.

BUT, IS IT FEASIBLE?



RESEARCH QUESTIONS

RQ1

How might **co-design** and **co-regulation** be defined, and why have they emerged as alternative approaches **in the context of AI governance**?

RQ2

In a **co-design model**, how can domain-specific expert knowledge, such as that of **legal professionals**, be effectively elicited and integrated to identify key dimensions and guiding principles for **responsible AI policies**?

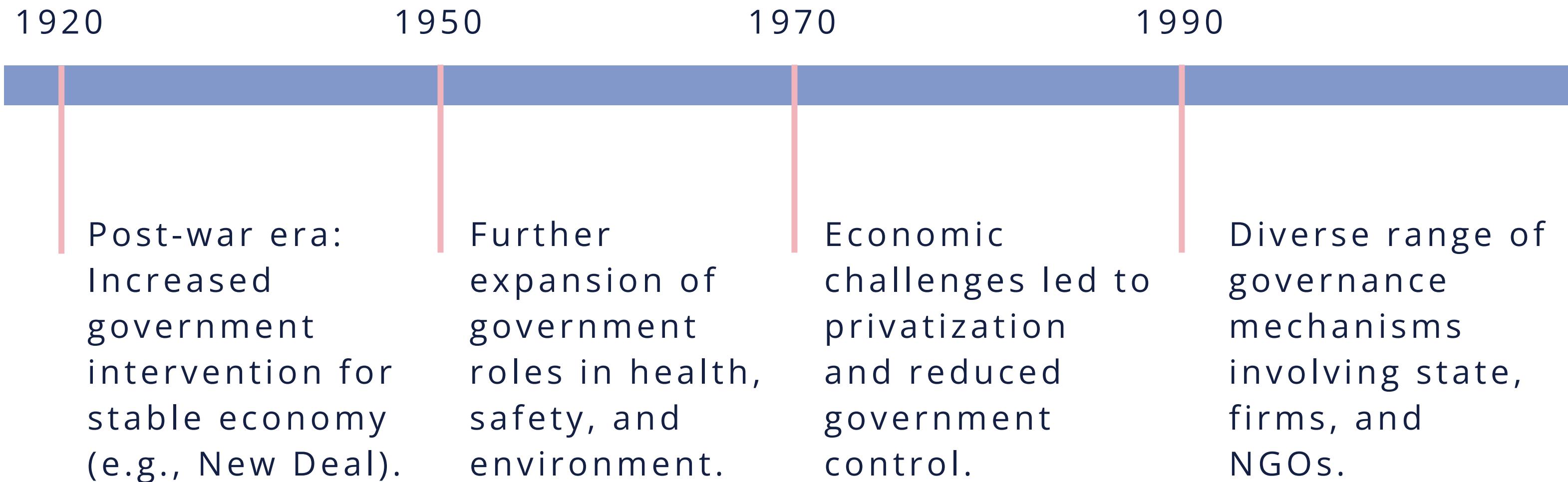
RQ3

How do **co-regulation models in related domains** facilitate collaboration among diverse stakeholders, and **what lessons** can be learned from these approaches to inform the development of **AI governance frameworks**?

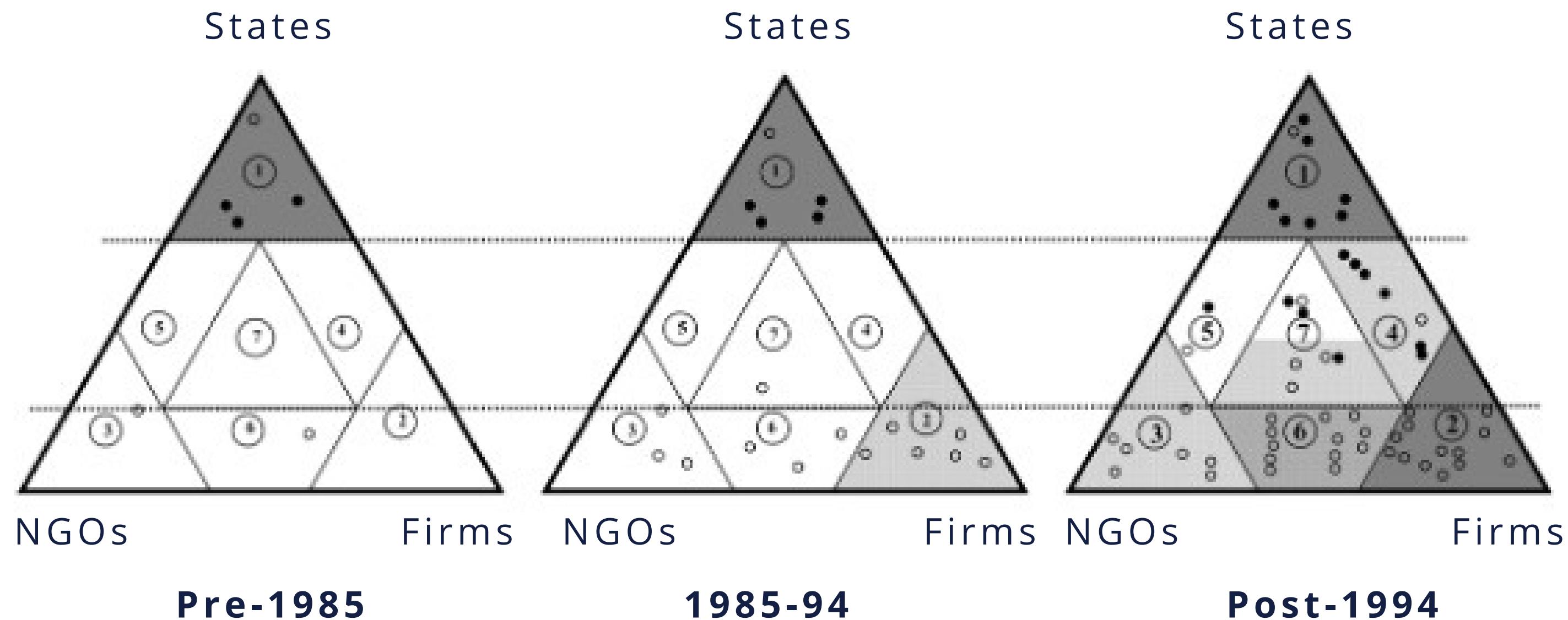
2

Theoretical Foundations

FROM GOVERNMENT CONTROL TO MULTI-STAKEHOLDER GOVERNANCE



GOVERNANCE TRIANGLES

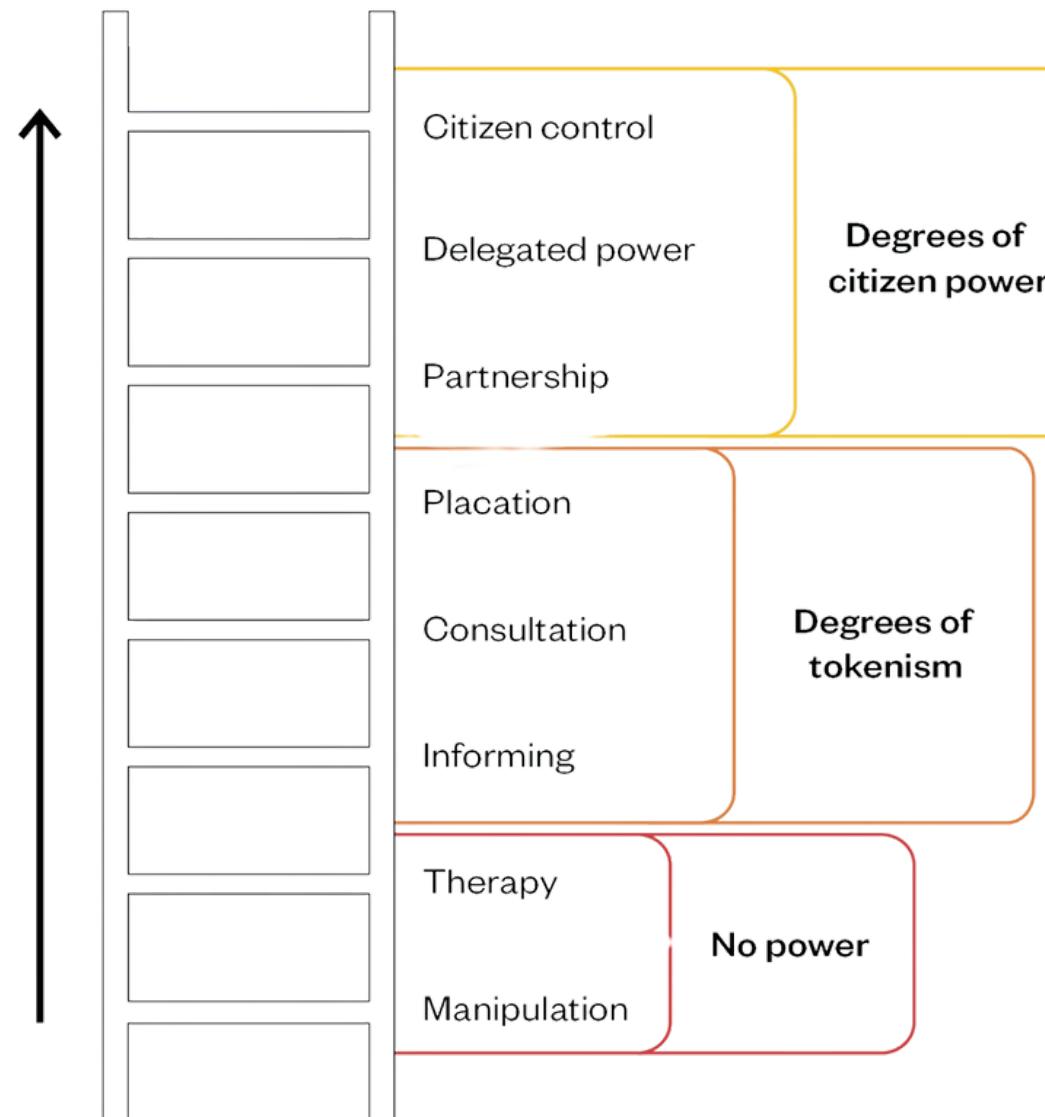


Kenneth W. Abbott and Duncan Snidal. The governance triangle: Regulatory standards institutions and the shadow of the state. In Walter Mattli and Ngaire Woods, editors, *The Politics of Global Regulation*, Princeton University Press, 2009.

AI REGULATION LANDSCAPE

COUNTRY	AI REGULATION	KEY FEATURES
EU	AI Act (2024)	<ul style="list-style-type: none">Bans certain AI uses (e.g., government social scoring)Risk assessment, transparency, registration, and human oversight requirements for high-risk AI
CANADA	AI and Data Act (Introduced in 2022)	<ul style="list-style-type: none">Risk assessment, transparency, registration, non-discrimination, and human oversight requirements for high-impact AI
CHINA	Interim Measures for Generative AI Services (2023)	<ul style="list-style-type: none">Prohibits generating content that endangers national security or social stabilityRisk assessment, license/approval, data protection, and non-discrimination requirements
US	Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI (2023)	<ul style="list-style-type: none">Tasks various federal agencies with specific responsibilities from cybersecurity to critical infrastructure to content authentication

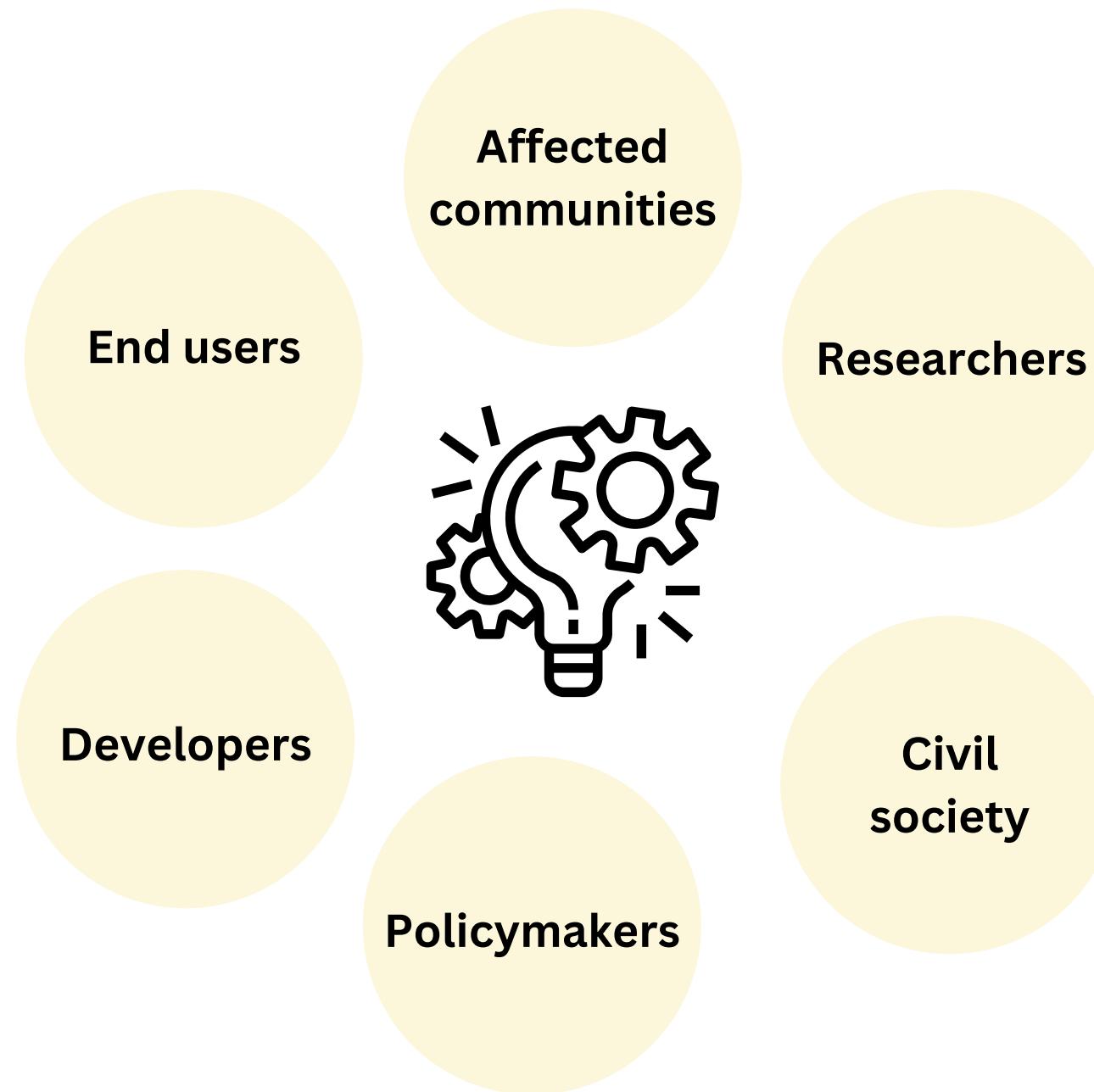
DEMOCRATIZING CENTRALIZED TECHNOLOGIES : PARTICIPATORY DESIGN (PD)



- Originated in Scandinavia in the 1970s and 80s.
- Workplace democracy → Urban planning, politics, technology design.
- Involves stakeholders, users, and communities in the design process.

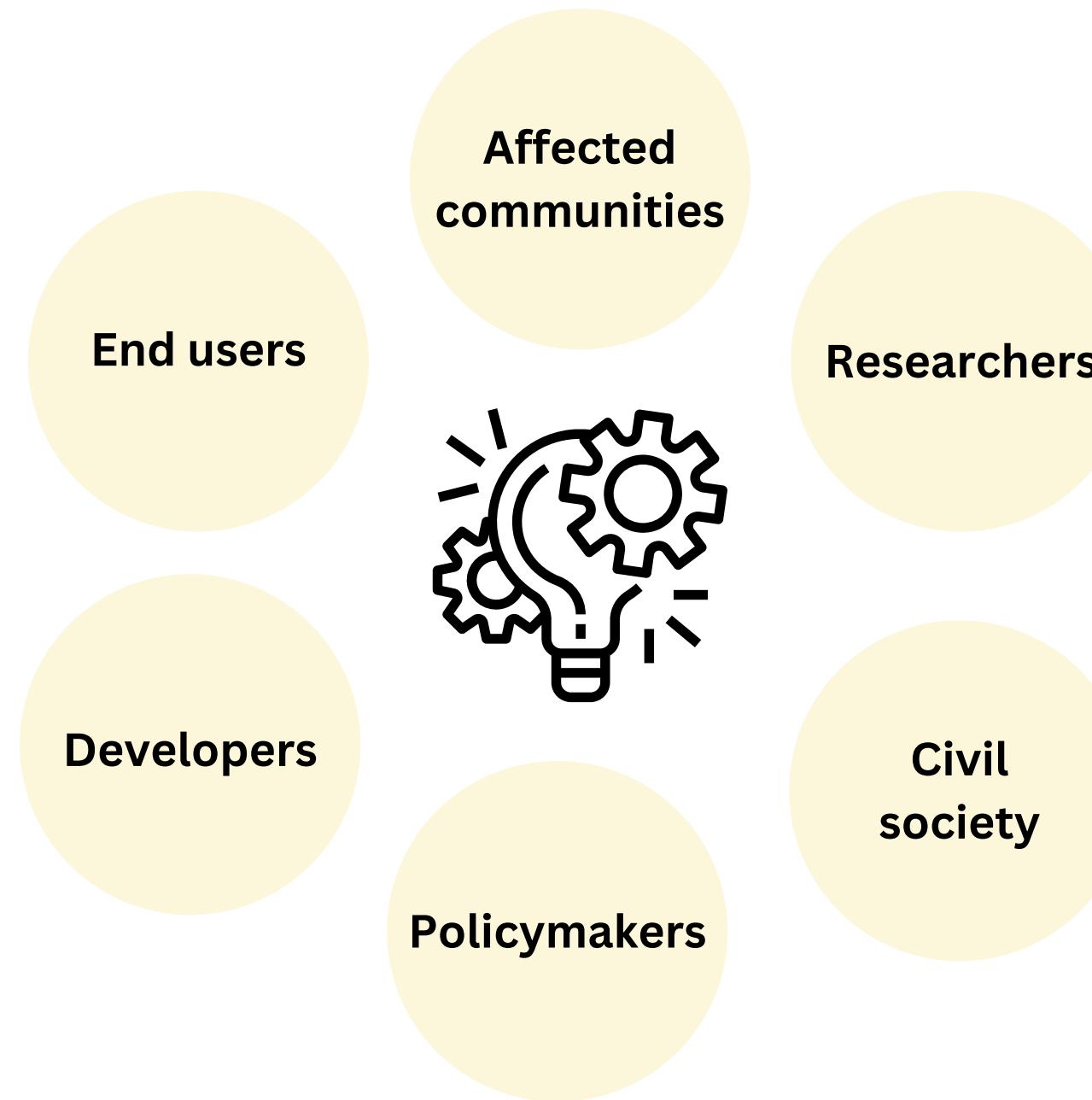
Arnstein's 'Ladder of Citizen Participation,' illustrated by Ada Lovelace Institute. Participatory data stewardship.
<https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/>, 2021.

CO-DESIGN



"A participatory process whereby relevant stakeholders, users, and affected communities in collaboratively developing a **particular socio-technical system**."

CO-REGULATION



"Collaborative development and **enforcement of universal rules and standards** by a range of stakeholders, including government agencies, industry representatives, civil society organizations, and experts."

ACROSS THE INDUSTRY

PUBLIC AUTHORITY

CO-DESIGN & CO-REGULATION

CO-REGULATION

CO-DESIGN

NATION

SUPRA-NATION

NATION

AI FIRM

AI FIRM

USER

USER

USER

USER

USER



3 Case Study 1 : Co-design in Legal AI

Inyoung Cheong, King Xia, K. J. Kevin Feng, Quan Ze Chen, and Amy X. Zhang. 2024. (A)I Am Not a Lawyer, But...: Engaging Legal Experts towards Responsible LLM Policies for Legal Advice. In The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24).

RESEARCH QUESTIONS

RQ1

How might **co-design** and **co-regulation** be defined, and why have they emerged as alternative approaches **in the context of AI governance**?

RQ2

In a **co-design model for AI governance**, how can domain-specific expert knowledge, such as that of **legal professionals**, be effectively elicited and integrated to identify key dimensions and guiding principles for **AI policies**?

RQ3

How do **co-regulation models in related domains** facilitate collaboration among diverse stakeholders, and **what lessons** can be learned from these approaches to inform the development of **AI governance frameworks**?

Methods



20 legal experts



Case-based reasoning

Workshop Procedure

Cases

1

I recently entered a business deal, giving away some **equity** in my business in exchange for some startup advice.... Will this work or **could I win** the resulting lawsuit?

2

I was brought in for questioning for a crime I did not commit. The officers employed a **lie detector test**... Is the lie detector's results **admissible** as evidence in court?

3

I run a popular online **mobile game** company... we collect user's data... run ads in our games, which also generate revenue. What should our **legal strategy** include, consider, or be?

...

[Randomly-chosen 20 cases are presented to participants.]

Templates

Content warning

This content may violate our content policy. If you believe this to be error...

Refuse

As an AI model, I am not a substitute for consulting with a qualified professional.

Empathize + refuse 

I am sorry that this is happening to you. It sounds like a challenging situation. Unfortunately, as an AI model...

Retrieve information

I can't give you legal advice specific to your situation but here is a list of laws and cases that may be helpful to look into...

Help work through questions 

In matters of divorce, legal proceeding often require addressing issues such as child custody, spousal support...

Help work through potential outcomes 

For charges related to intoxicated driving, the penalties can range from A to B based on the severity...

Recommend actions

As a prudent next step, I recommend initiating formal proceedings by lodging a complaint with the...

Workspace



[Expert's Name]

[Case 1] [I run a popular online mobile game...]

Appropriate Response

Refuse + empathize,
Help work through questions

Dimensions

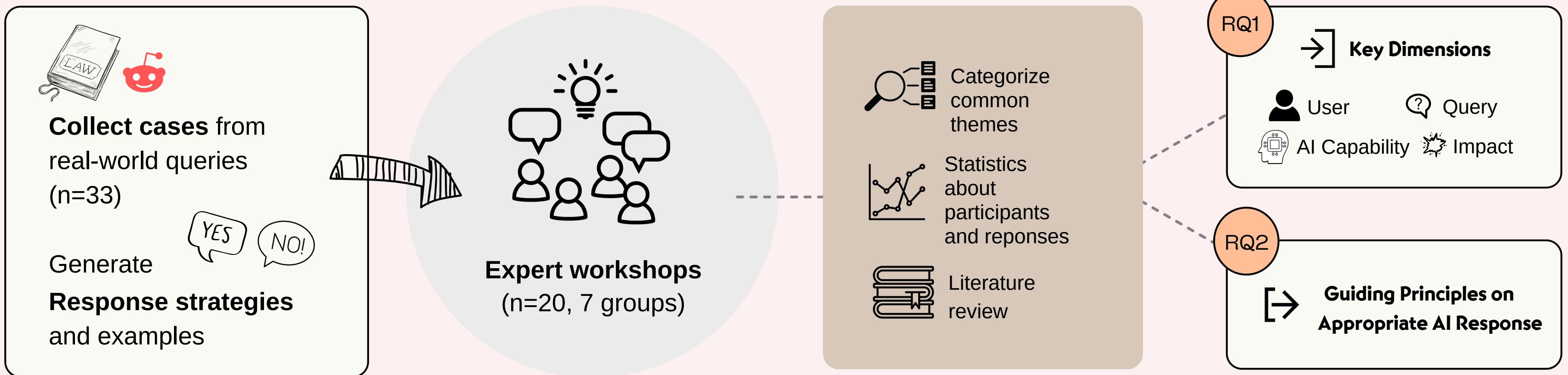
- Information privacy
- Jurisdictions (Which state? Other countries?)
- Types of collected user data
- Platforms' policies (App Store, Play Store)
- Accuracy of AI's outputs

[Case 2] [My husband and I have been married...]

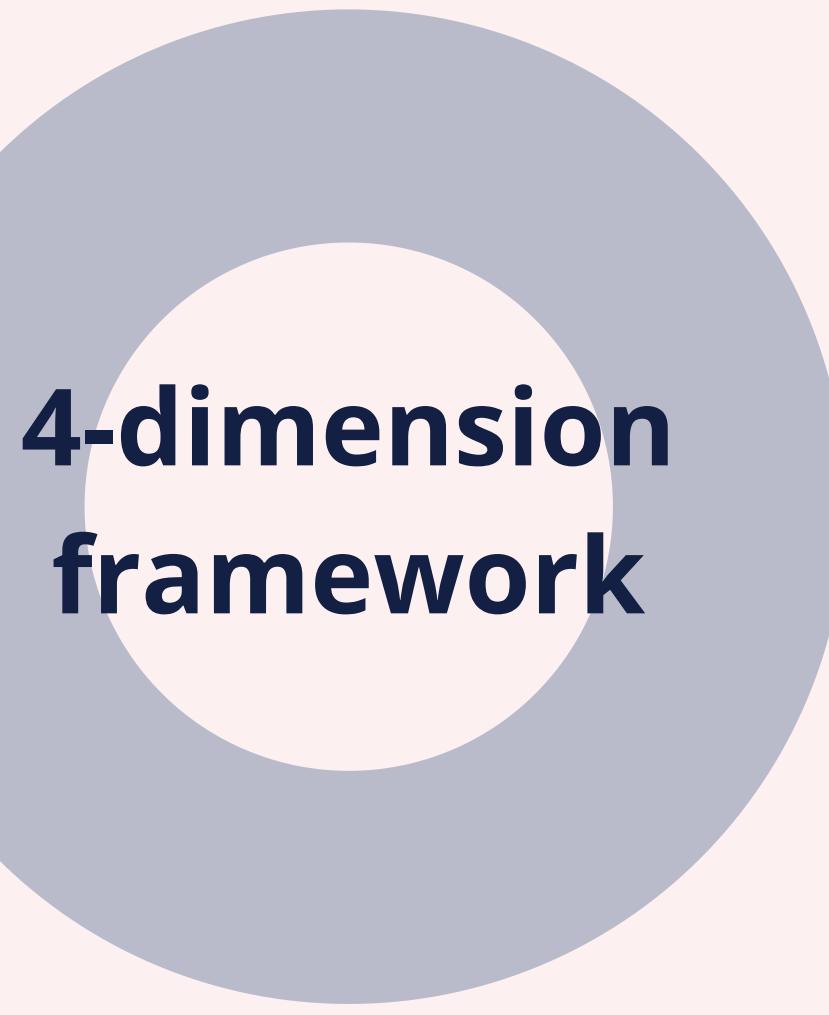
...

[Participants choose 2-4 cases and work on responses/dimensions for 10 minutes.]

Research Process

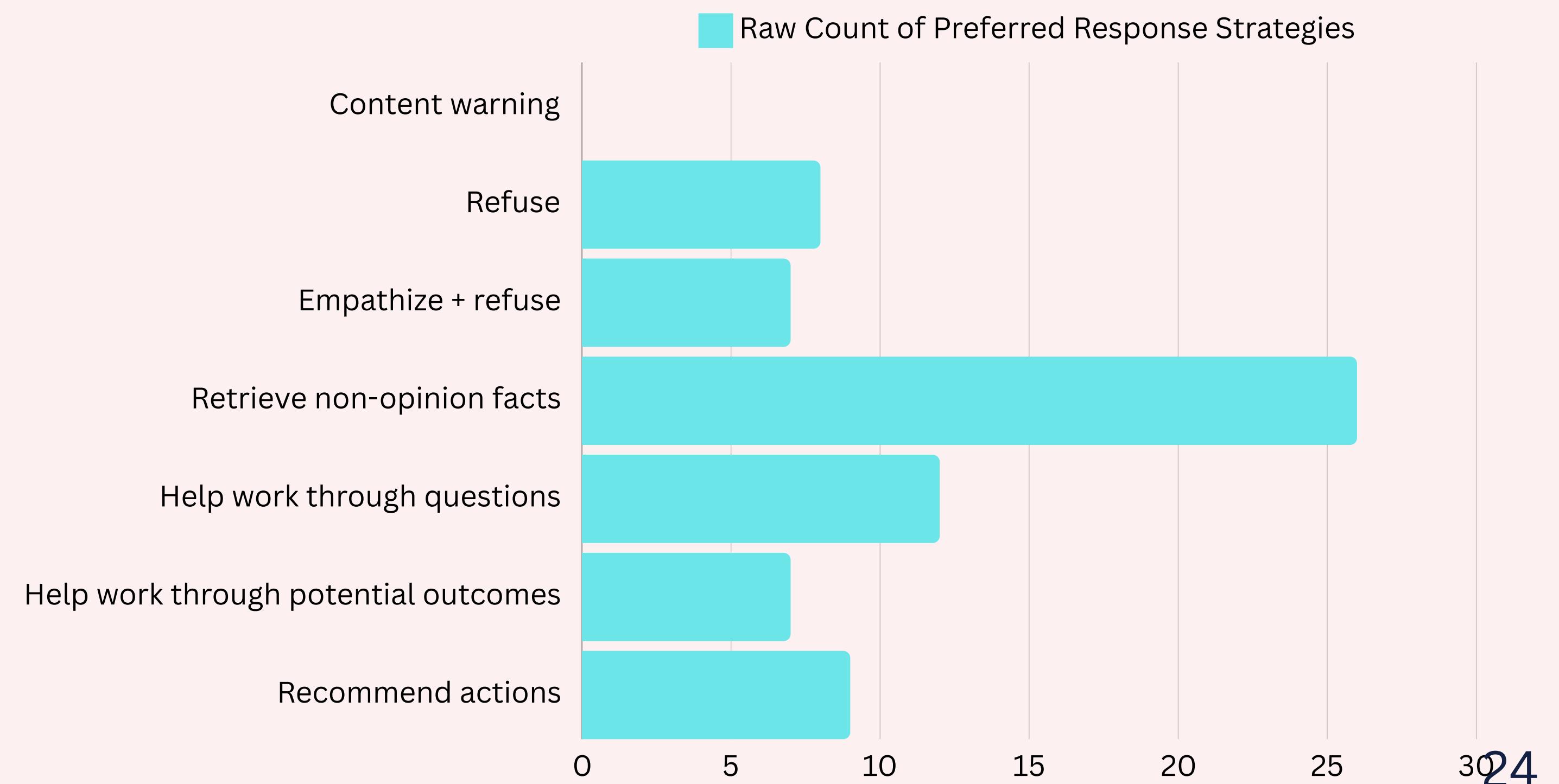
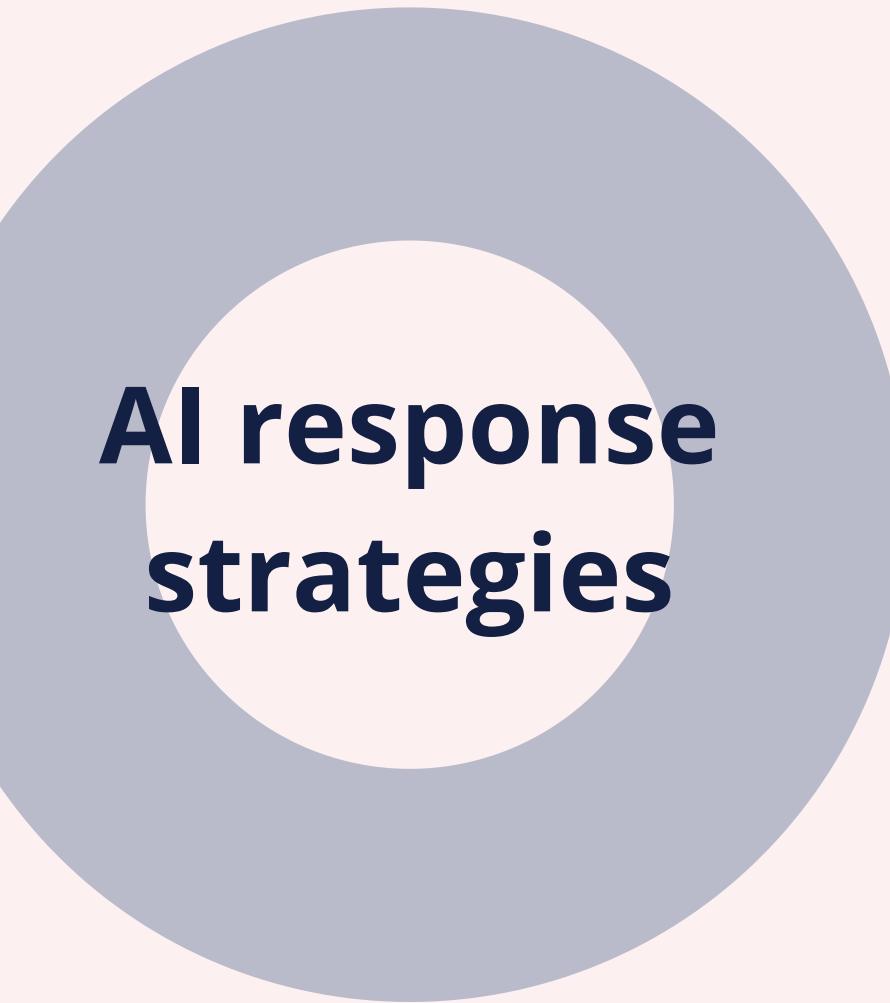


Results

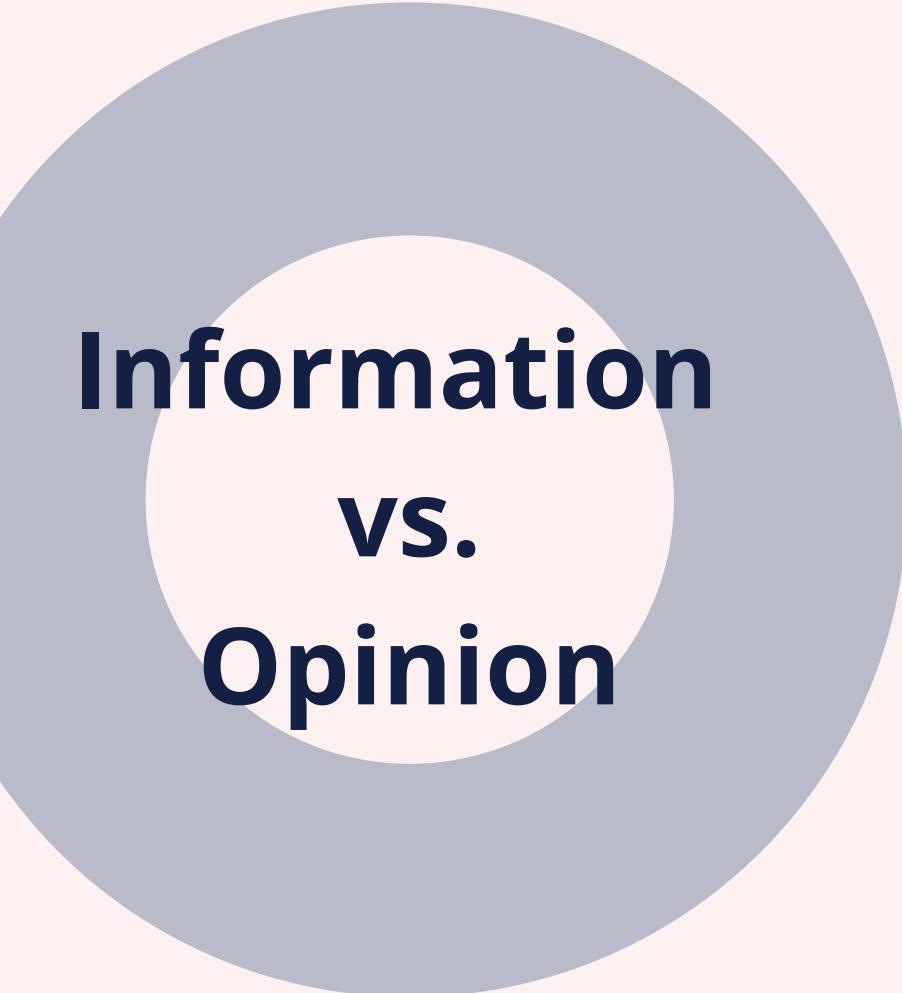


USER	ATTRIBUTE	BEHAVIOR	
	<ul style="list-style-type: none">• Identity (age, ethnicity, gender...)• Geography• Vulnerable/marginalized group• Sophistication• Available resources	<ul style="list-style-type: none">• Ambiguity• Intent• Reliability• Agency	
QUERY	FACT	LAW	DESIRED ANSWER
	<ul style="list-style-type: none">• Existing contracts• Third-parties• Stakes	<ul style="list-style-type: none">• Complexity• Criminal v. Civil• Subject matters (tax, privacy, Constitution)	<ul style="list-style-type: none">• Information• Opinion• Other advice (emotional support, ethical/practical guidance)
AI	<ul style="list-style-type: none">• Accuracy• Context-awareness• Confidentiality• Accountability• Bias		IMPACT <ul style="list-style-type: none">• Impact on user (emotional, cultural, practical)• Impact on others

Results



Results



Information
vs.
Opinion

Do not make definitive judgement about specific cases.

Do not predict outcomes.

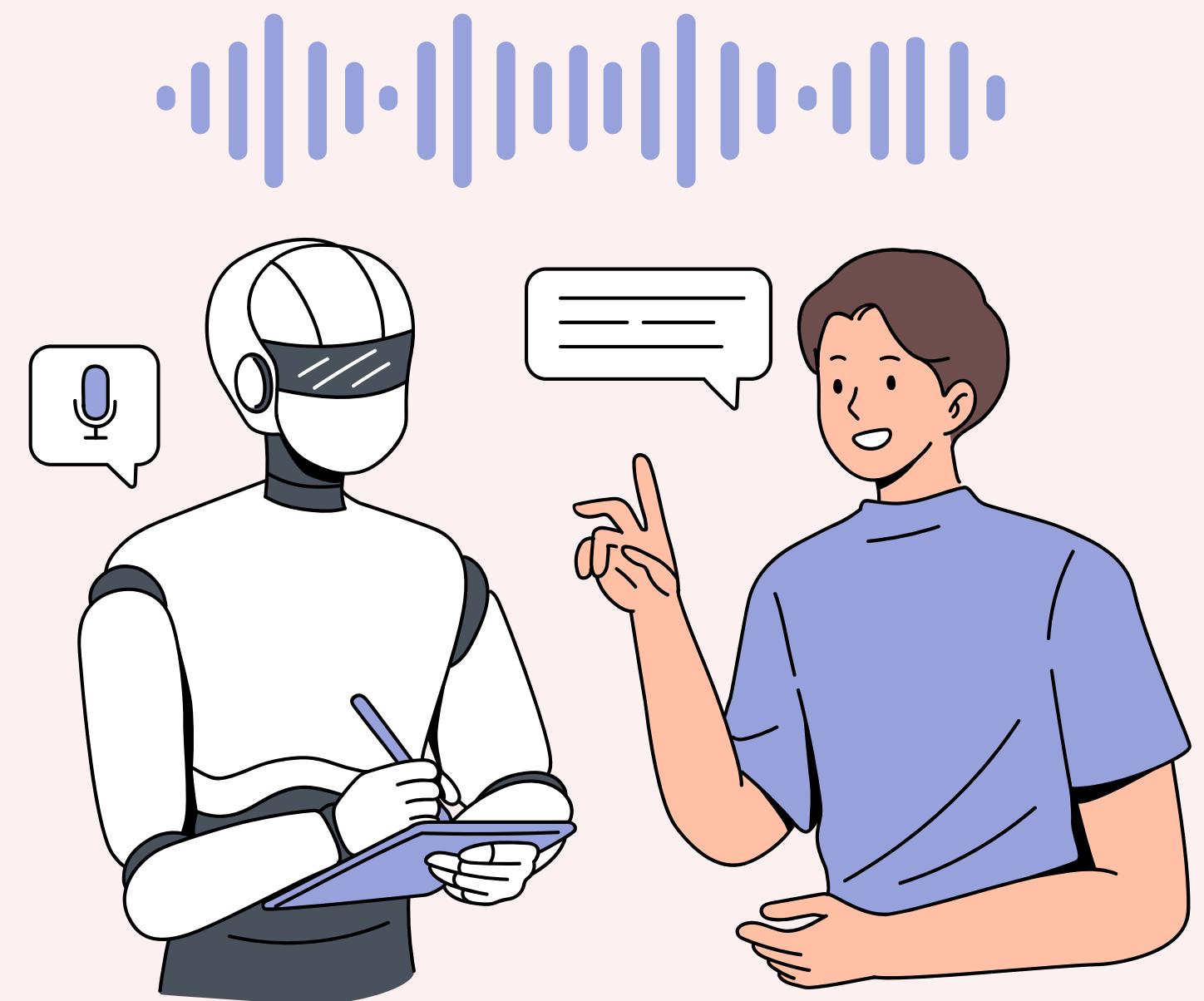
Do not recommend actions.

Do not provide a cost-benefit analysis.

Results

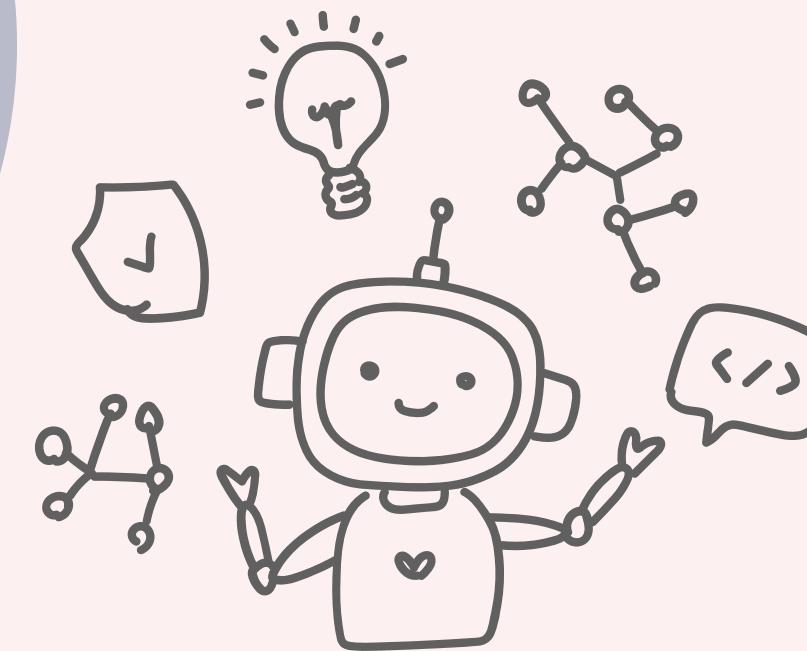
Multi-turn
Interactions

- Elicit relevant facts
- Refine questions
- Point to helpful resources
for further research



Results

Other
Guiding
Principles



- Don't pretend to be human.
- Caveat constraints.
- Avoid potential harm.
- Respect the justice system.
- Avoid unethical answers.
- Be transparent.
- Avoid appearance of impropriety.

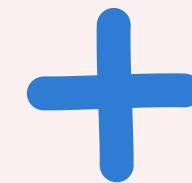
Discussion

Benefits of case-based reasoning methods



Clinical experience
Professional knowledge

Query-specific concerns



Overarching
legal/technical concerns

Discussion

Charting novel legal considerations

- Users lack confidentiality and accountability protections governing attorney advice.
- Unauthorized practice of law regulations prohibit non-lawyers from legal advising, carrying criminal penalties.

Discussion

Learning from time-tested wisdom

Type	Permissible questions	Impermissible questions
Procedure	Can you tell me how to file a small claims action?	Can you tell me whether it would be better to file a small claims action or a civil action?
Definition	What does "certificate of service" mean?	My neighbors leave their kids at home all day without supervision. Isn't that child neglect?
Forms	I need to file for divorce and I have no idea where to begin. Is there some place I can go to find out how to get started?	The self-help divorce petition says I should list any gifts as my separate property. Should I list the money that my parents gave me last month as my separate property?
Options	What can I do if I cannot afford to pay the filing fee?	My ex-husband hasn't paid the debts that he agreed to pay in our divorce settlement. Can I be made responsible for this debt?

Discussion

Applicability to Other Professional Domains



4-dimension
framework

User ? Query
AI Capability Impact



CONTRIBUTIONS & LIMITATIONS

EXAMPLE OF CO-DESIGN IN PRACTICE

Illustrates how expert knowledge can be integrated into AI policy discussions.

Showcases successful gathering of legal professionals' perspectives under time constraints.

GROUNDED AND NUANCED UNDERSTANDING

Achieves detailed, context-specific insights through workshop-based interactions, moving beyond abstract, high-level discussions.

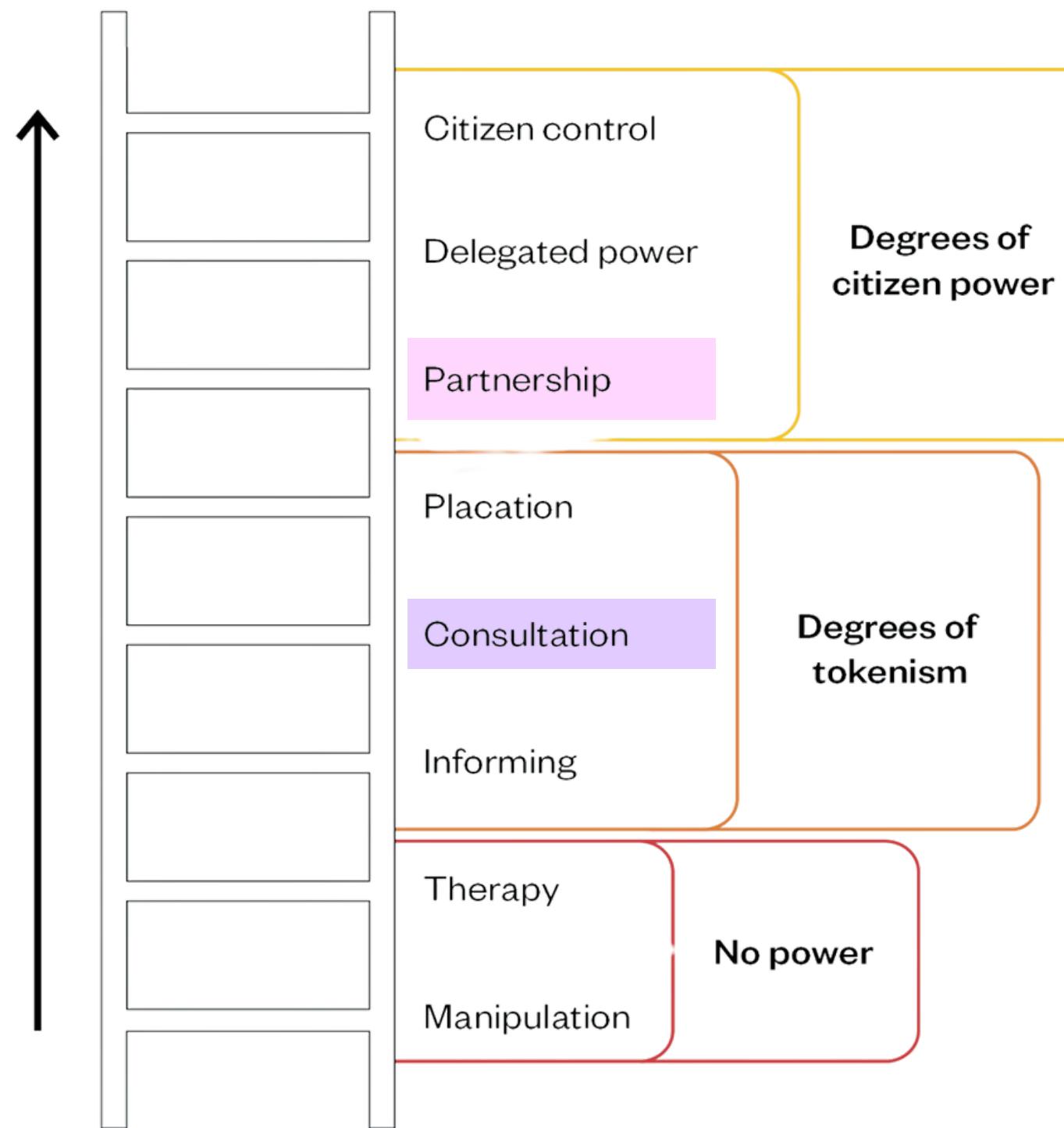
Reveals nuanced critiques and considerations grounded in professional experience.

LEVEL OF PARTICIPATION ACHIEVED

Reaches the "consultation" level but falls short of full "partnership" level on Arnstein's ladder of participation.

Needs to separately examine end users' benefits and costs of using AI for legal advice.

Further Research



Case Law for AI Policy

Democratically Aligning AI with Human Preferences through Case Repositories

Ja					AFFILIATIONS University of Washington
					Independent Attorney

<https://social.cs.washington.edu/case-law-ai-policy/>



4

Case Study 2 : Co-regulation of Online Content in South Korea

w/ Pardis Emami-Naeini and Tadayoshi Kohno as advisors

RESEARCH QUESTIONS

RQ1

How might **co-design** and **co-regulation** be defined, and why have they emerged as alternative governance approaches **in the context of AI governance**?

RQ2

In a **co-design model for AI governance**, how can domain-specific expert knowledge, such as that of **legal professionals**, be effectively elicited and integrated to identify key dimensions and guiding principles for **AI policies**?

RQ3

How do **co-regulation models in related domains** facilitate collaboration among diverse stakeholders, and **what lessons** can be learned from these approaches to inform the development of **AI governance frameworks**?



SOUTH KOREA

DEMOCRATIZED IN THE 1990S

Freedom of Expression barely existed under the monarchy, Japanese colonization, and military dictatorship.

COLLECTIVIST CULTURE

Korea Constitution Article 21(4) "Neither speech nor the press shall violate the honor or rights of other persons nor undermine public morals or socialethics. Should speech or the press violate the honor or rights of other persons, claims may be made for the damage resulting therefrom."

COURTS UPHOLDING REGULATION ON ONLINE CONTENT

Despite the abolition of censorship laws, courts upheld government agencies' take-down authorities over illegal online content.

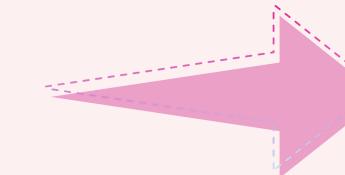
SOUTH KOREA



Creators resist against
government direct regulation.

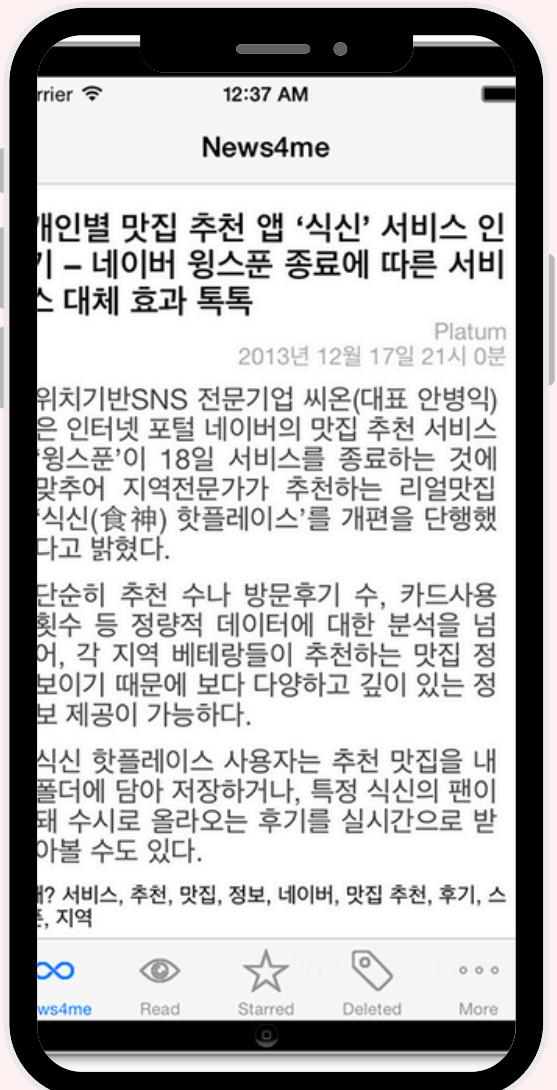


Major online platforms want to
make peace with creators,
governments, and civil society.

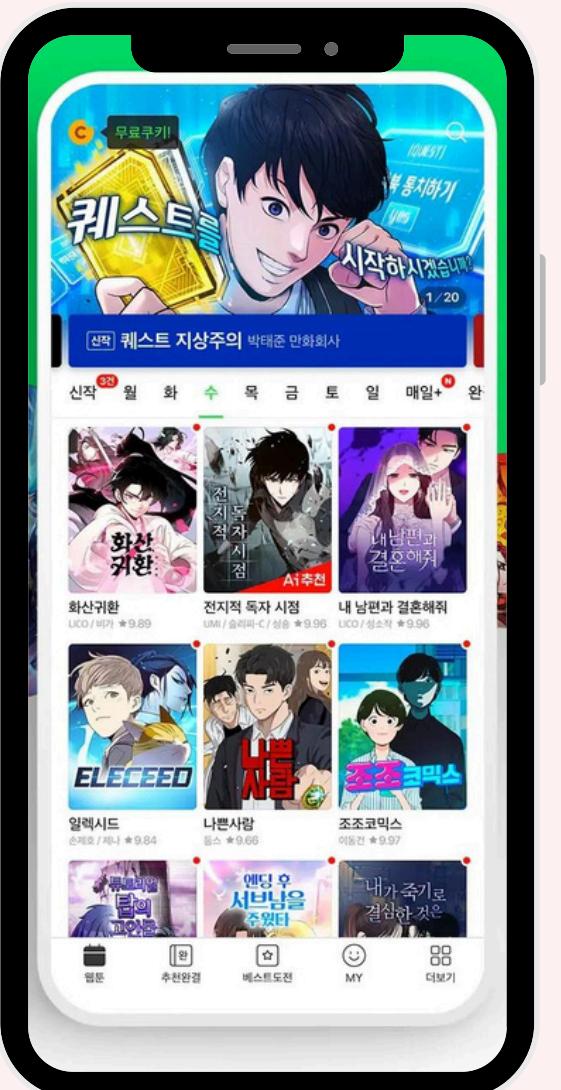


SOUTH KOREA

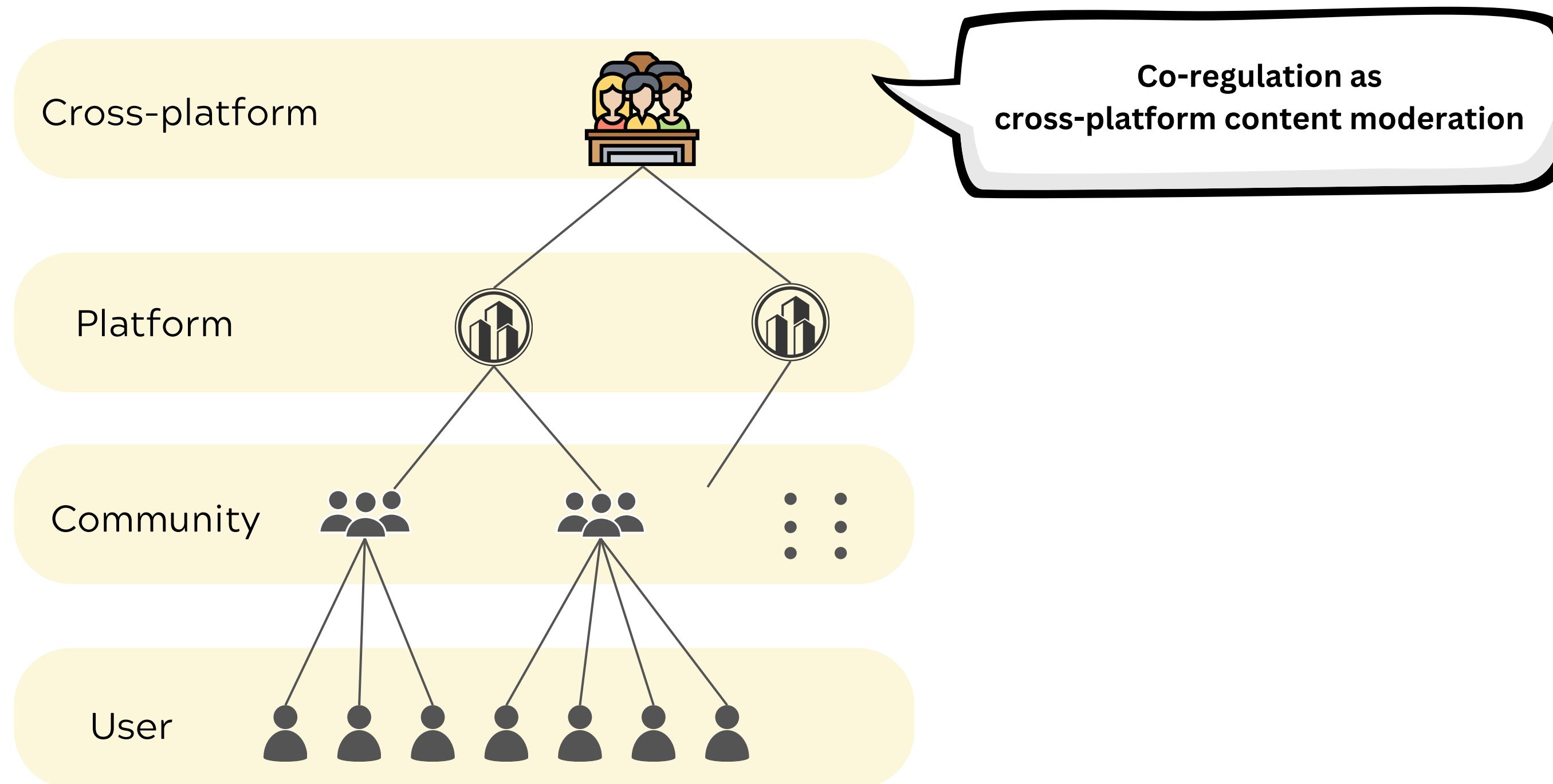
News



Comics



TYPOLOGY OF CONTENT MODERATION

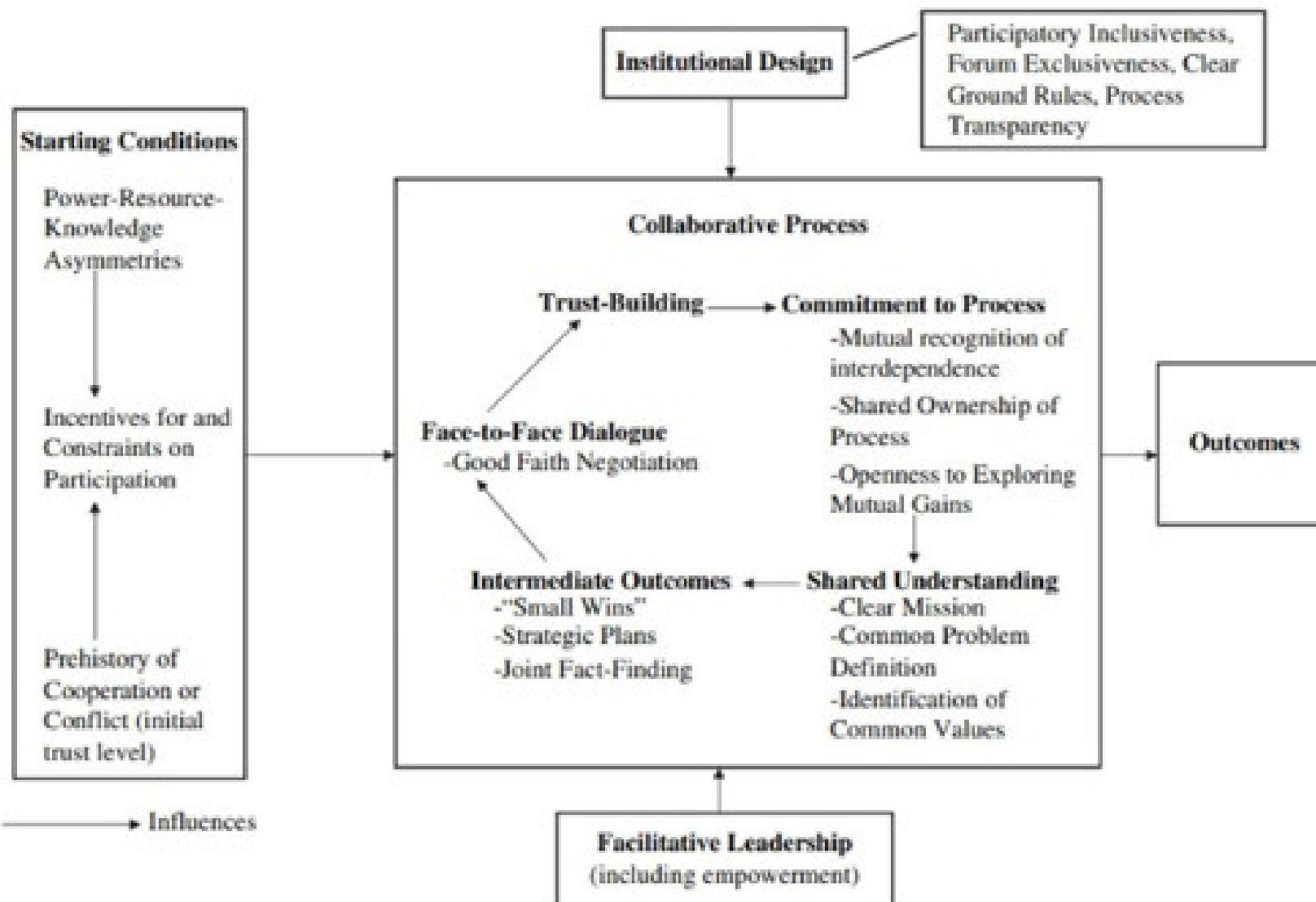


METHODS

- In-depth interviews with 15 key stakeholders in 2022
 - 4 creators
 - 4 co-regulatory board members
 - 4 platform employees
 - 3 government actors
- Field observations in Fall 2022
- Document analysis
 - Annual reports, meeting notes
 - Legislative documents, rules and regulations
 - News articles
- Comparative historical analysis

ANALYTICAL FRAMEWORK

ANSELL & GASH'S COLLABORATIVE GOVERNANCE MODEL



Chris Ansell and Alison Gash. Collaborative governance in theory and practice. Journal of public administration research and theory, 18(4):543–571, 2008.

- **Starting conditions:** These include power-resource-knowledge asymmetries, incentives for and constraints on participation, and the prehistory of cooperation or conflict.
- **Institutional design:** This encompasses participatory inclusiveness, forum exclusiveness, clear ground rules, and process transparency.
- **Facilitative leadership:** Strong leadership is crucial in building trust, facilitating dialogue, and exploring mutual gains.
- **Collaborative process:** This involves face-to-face dialogue, trust building, commitment to the process, shared understanding, and intermediate outcomes.

STARTING CONDITIONS

INTER-
DEPENDENCE

PRE-HISTORY
OF
ANTAGONISM/
COOPERATION

INCENTIVE TO
PARTICIPATE

NEWS

- News organizations work independently
- Platforms can't change news content

- Distrust and resentment toward platforms
- Feel platforms took away their direct audience connection

- Courts seen as more effective than co-regulation
- No government backing ("shadow of the state")

COMICS

- Artists rely heavily on platforms
- Platforms handle distribution and support production

- See platforms as helpful partners
- Appreciate new business opportunities from platforms

- Co-regulatory body (ACW) is the main decision-maker
- Government can step in if co-regulation fails

INSTITUTIONAL DESIGN

INCLUSIVE-NESS

- Boards include journalists and journalism professors

FORUM EXCLUSIVENESS

- Many alternative forums exist (courts, Press Arbitration Commission)
- Co-regulatory bodies (PEC, INC) are not the only option
- Broad, general guidelines on journalistic principles

CLEAR GROUND RULES

- Board includes critics, professors, and experts
- Creators are not directly represented

- ACW has exclusive authority over web comics ratings
- Legal challenges are rare

TRANS-PARENCEY

- Decision outcomes are publicly disclosed. Decision-making process is not fully disclosed because of fear of public scrutiny discouraging participation and risk of external pressure on board members
- Some stakeholders advocate for more transparency

NEWS

COMICS

- Clear checklists for issues like violence and sexual content
- Platforms find these rules practical and easy to apply

FACILITATIVE LEADERSHIP

NEWS

- Few successful collaborations between media companies
- Competitive mindsets hinder collective action
- Journalists skeptical of co-regulatory decisions made by peers
- Online platforms Maintain distance from news content regulation
 - They view themselves as mediators, not publishers

COMICS

- Cohesive environment among creators. Trust in peers to represent community interests
- Korean Comics Artists Association (KCAA) plays a leadership role
- "Organic leaders" in creator communities, expert groups, and platforms
- Major platforms (Naver and Kakao) committed long-term support and funding

LEADERSHIP VACUUM

COLLABORATIVE MINDSET

COLLABORATIVE PROCESS

NEWS

- Despite co-regulatory bodies' efforts, participants view trust-building activities as naive

- Journalists' sense of independence can make them resistant to external influence

- Struggles to achieve common vision due to fragmented industry
- Difficulty in evaluating harm in news content (e.g., misinformation)

- Some small wins in specific areas (e.g., reporting on suicides, election polls)

COMICS

- Spent time clarifying expectations and roles before official launch
- Positioned as a partner, not a regulators

- Despite the voluntary nature, higher and stable commitment from stakeholders

- Easier to evaluate and agree on content issues (e.g., age-appropriate content)

- Achieved industry-wide standards and content ratings
- Successes widely promoted among stakeholders

TRUST-BUILDING

COMMITMENT

SHARED UNDERSTANDING

INTERMEDIATE OUTCOMES

CONFIRMING UTILITY OF FRAMEWORK

COMICS

- History of cooperation with online platforms
- High interdependence between creators and platforms
- Strong facilitative leadership from "organic leaders"
- Exclusive authority of ACW for content-related decisions
- Clear and detailed ground rules, especially for content rating
- Shared understanding among stakeholders about goals and balance

NEWS

- Low trust among stakeholders and towards co-regulatory bodies
- Weak incentives for participation due to alternative forums
- Fragmented industry with many media companies
- Strong sense of journalistic independence resisting oversight
- Difficulty in achieving consensus on complex issues like misinformation

COMPLEMENTING THE FRAMEWORK

**ANSELL
& GASH**

- Address power balances
- Ensure high independence among stakeholders
- Make the collaborative forum an exclusive venue
- Address prehistory of antagonism through trust building
- Rely on strong “organic leaders” when power is asymmetrical...

**NOVEL
PRINCIPLES**

- **Legal and Normative Considerations:** Consider impacts of free speech and anti-competition laws for online content regulation.
- **Funding Stability and Independence:** Ensure basic operational funding is not threatened by participant exits. Maintain longevity and effectiveness.
- **Enforcement and Monitoring Mechanisms:** Establish clear signals for compliant and non-compliant actors. Clarify incentive structures to encourage consistent participation

CHALLENGES IN AI GOVERNANCE

NORMATIVE CHALLENGES

- Tensions between co-regulation & free speech principles
- Collaboration on safety standards could be seen as anti-competitive

PRACTICAL CHALLENGES

- Unclear incentive to participation
 - Lack of enforcement can lead to non-compliance
 - Free-rider problem when benefits are intangible or long-term
 - Difficulty in defining and measuring compliance in AI
- Participatory inclusiveness
 - Difficulty in ensuring meaningful participation from all stakeholders given the AI's general-purpose nature
 - Challenges in addressing diversity in closed-source, open-source, and various AI applications as well as different legal/cultural landscapes
- Stable funding and independence
 - Significant costs for expertise, monitoring, and global operations
 - Industry funding risks regulatory capture; Government funding may invite political interference; Donation-based funding likely insufficient for AI governance scale

GUIDING PRINCIPLES

STARTING CONDITIONS

Acknowledge power disparities

Leverage interdependence among stakeholders

Be aware of existing conflicts/cooperation

Navigate normative challenges (free speech/ antitrust)

INSTITUTIONAL DESIGN

Segment governance into themes (e.g., labor, creativity, professional services) for context-specific regulations

Consider officially creating exclusive forum

Aim for diversified funding sources

Monitor and sanction non-compliance

COLLABORATIVE PROCESS

Facilitate face-to-face dialogue

Budget time and resources for trust-building, esp. topics with prior conflicts

Focus on “small wins” such as rules for specific contexts

Ensure capacity to provide technical support for under-resourced members

FACILITATIVE LEADERSHIP

Identify and cultivate “organic leaders”

Leaders should articulate the collective benefits of participation

Encourage leaders to champion ethical AI development practices

5

Paths Forward

LESSONS LEARNED

IMPORTANCE OF CONTEXT

- Industry-specific dynamics, stakeholder relationships, and historical contexts significantly shape co-governance success
- AI governance in professional domains (e.g., legal advice) requires tailored policies incorporating input from diverse stakeholders

BARRIERS TO AI CO-GOVERNANCE

- Fragility of collaborative structures due to their flexible and voluntary nature
- "Participatory ceiling" in generative AI systems due to corporate control and context-agnostic nature
- Challenges in aligning diverse interests from corporations, government, and civil society
- Tension between content regulation and free speech
- Competing demands for transparency and data protection
- Global nature of AI development leading to regulatory fragmentation

GUIDING PRINCIPLES FOR AI CO-GOVERNANCE

FOCUS ON CONTEXT-SPECIFICITY

- Tailor approaches to specific contexts and stakeholder needs
- Incorporate field-specific wisdom and established norms
- Define scope, audience, and jurisdiction of co-governance efforts

GOVERNING AI AS A HUMAN-CENTRIC PROCESS

- Utilize human-centric methodologies (e.g., case-based deliberation, confidential interviews)
- Prioritize sustained, well-designed processes for meaningful stakeholder engagement
- Build trust among stakeholders through ongoing, iterative processes

RESOLVING LEGAL AMBIGUITIES

- Address normative uncertainties crucial for effective AI governance
- Balance free speech considerations with the need for regulation
- Draw parallels with existing regulatory frameworks (e.g., IRB ethics regulation)

CONTRIBUTIONS

- Offers strategic governance considerations that are crucial for navigating the evolving landscape of AI co-design and co-regulation
- Provides a foundation for understanding context-specific challenges in AI governance
- Sets a precedent for future research to build upon, addressing broader stakeholder perspectives and more general AI applications

LIMITATIONS

LIMITED FOCUS ON SPECIFIC DOMAINS

- Concentrated on AI providing legal advice, news apps, and comics apps
- General-purpose AI systems can be applied in countless unforeseen ways
- Governance models developed for specific domains might face challenges when scaled to broader AI applications

STAKEHOLDER ENGAGEMENT

- Primarily engaged with experts and professionals (creators, reporters, online platforms). Lacks direct input from end users' perspective
- Rapid evolution of AI capabilities constantly reshapes stakeholder landscapes. Shifting power dynamics among stakeholders as AI impacts various sectors differently.

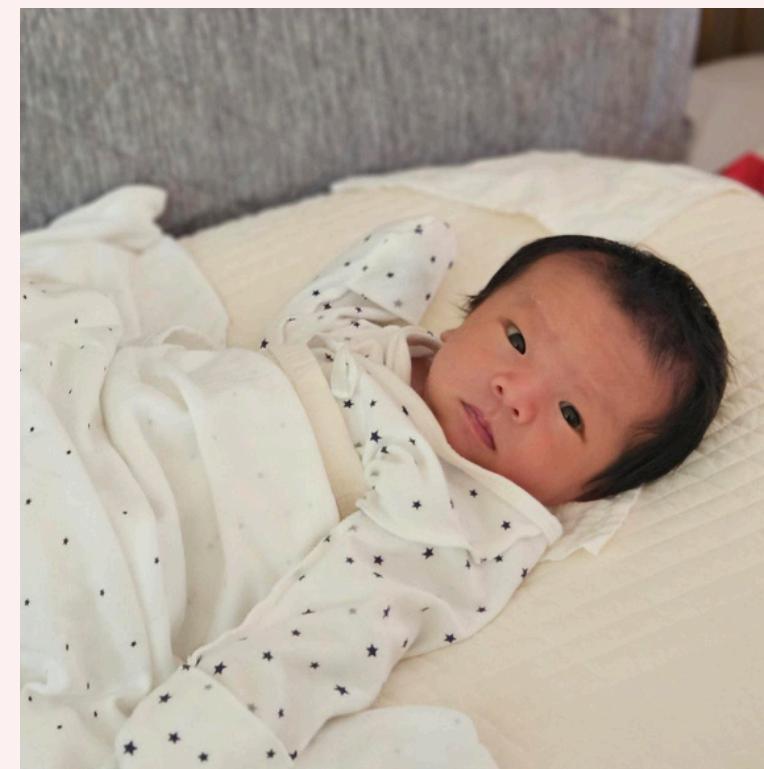
POLICY IMPLEMENTATION CHALLENGES

- Need to explore political and economic barriers to adopting co-design and co-regulation approaches

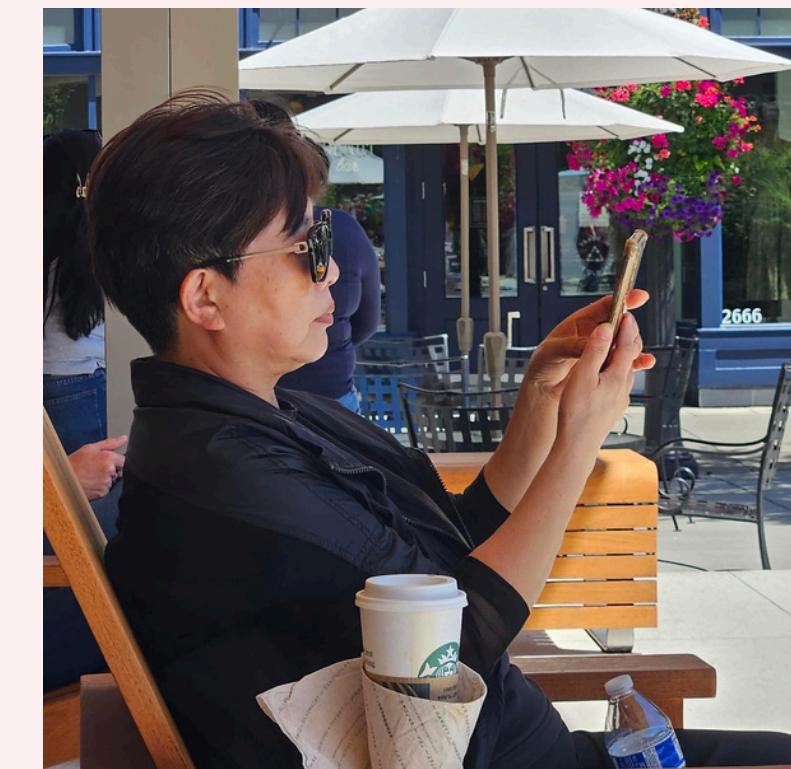
SPECIAL THANKS TO



KYOUNGCHE RAY KIM



ROY YUEL KIM



JINYOUNG BAE



THANK YOU
