Harnessing Generative AI

Arisa EMA
The University of Tokyo

Discussion Point (1)
AI and Risk

1. Risks posed by AI
   ① Risks arising from AI (machine learning) specific features
      - Uncontrollability, reduced transparency, etc.
   ② Risks common to information technology
      - Privacy, security, data bias, misuse, wrong answers, confidentiality leaks, risk of infringing copyright or portrait rights, etc.

2. Obstacles to AI Promotion
   ① The current system is an obstacle to the design, development, operation, use of AI, and the relationship between the application of AI services is unclear and innovation is inhibited
      - Business laws such as the Medical Practitioners Act, copyright laws, personal information protection laws, road traffic laws, etc.

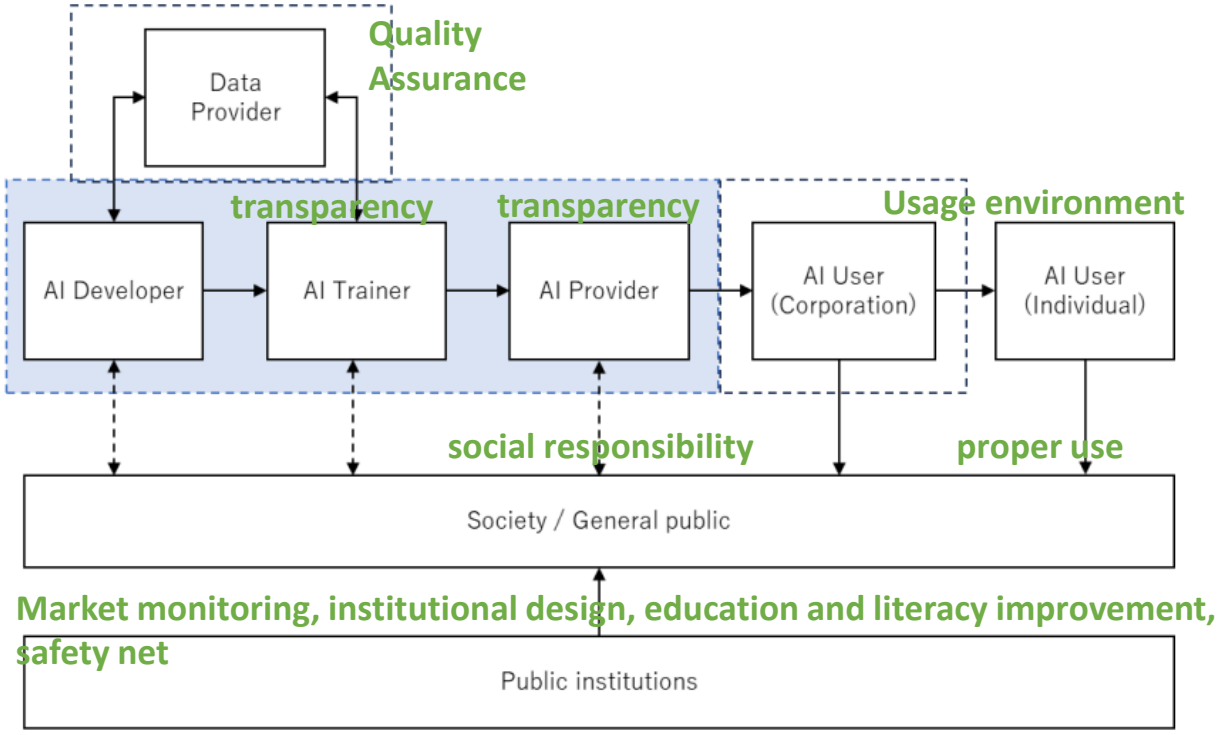
Discussion Point (2) Interoperability

- “Interoperability among frameworks" was the keyword for AI governance in the communiqué of the G7 Hiroshima Summit

- What is interoperability?
  1. international standard
     - Standardization of ISO, IEEE, NIST, CEN-CENELEC, etc.

  2. Interoperability between frameworks
     - Ensure interoperability (transparency) of "frameworks" for risk identification, assessment, response, so that disciplines for AI in different countries, regions, and organizations can coexist and cooperate.

Discussion Point (3)
AI System Stakeholders and Responsibilities



■Between businesses (in blue box)

- Agreements and commitments are exchanged between **businesses** to address risks and take responsibility.
- **Public** agencies monitor fairness of contracts, etc.

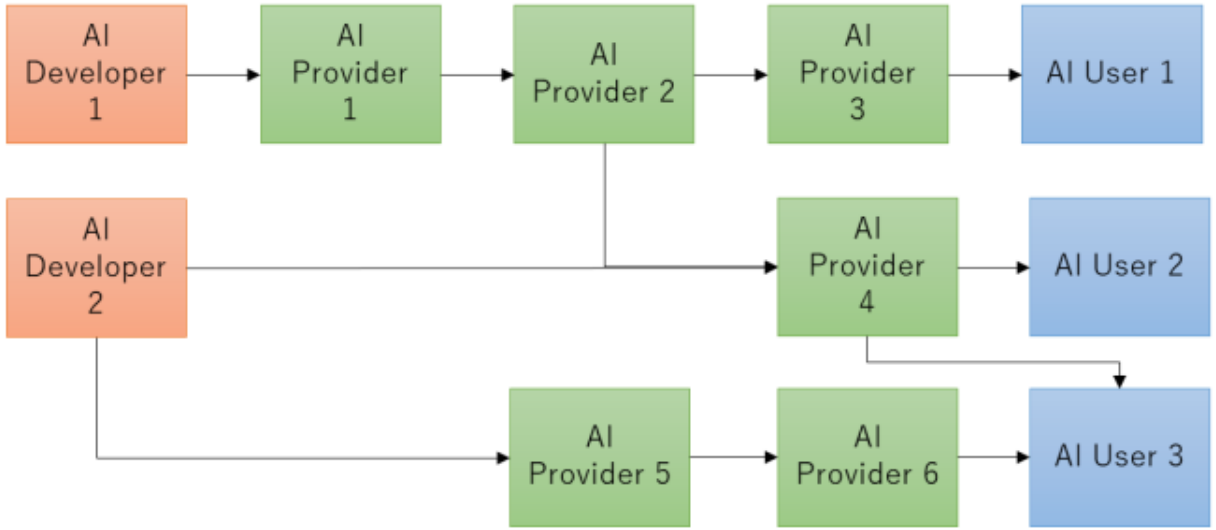
Business to consumer

- **AI provider** provides pre- and post-response
- **Users** also make proper use of the system.
- **Public institutions also** form a mechanism to investigate the causes of incidents and accidents, and to provide relief to victims.

AI development and deployment across organizations and countries

(1) Example of the same organization being responsible for AI development to provision

(2) Examples of complex and long supply chains where different organizations are responsible for AI development to provision



Interaction between Society and Technology

- AI/Robots that ***solve*** the social problems?
- AI/Robots that ***create*** the social problems?

What kind of society do we want to live in?
And how to design the society and technologies?

We need to have a conceptual investigation:
unpack underlying assumptions of the concept
to start cross-cultural, inter-disciplinary dialogue

Global AI Dialogue

ABOUT THE DIALOGUES

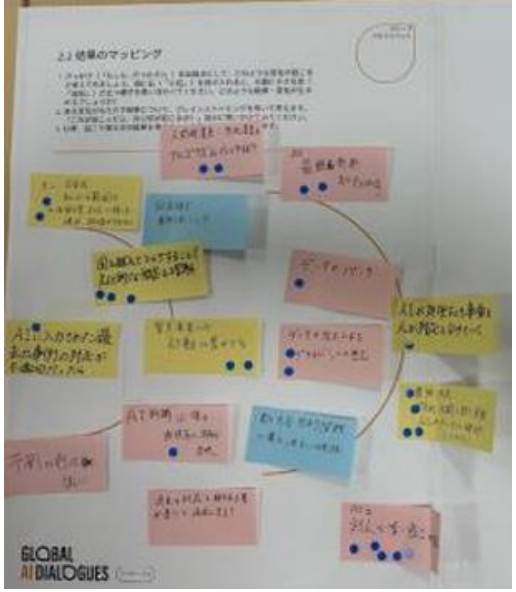
MUNICH, GERMANY
Prof. Jens Grossklags, Chiara Ullstein, Michel Hohendanner

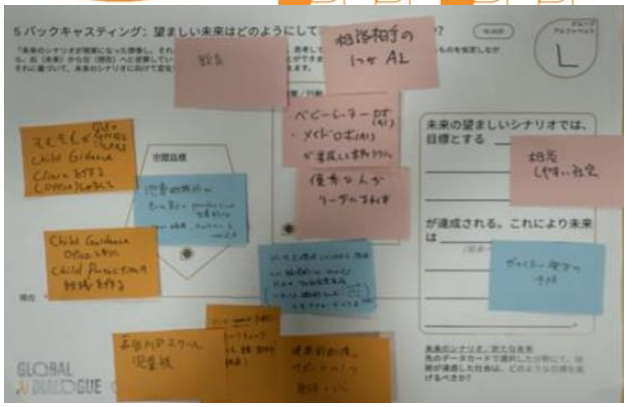
LAGOS, NIGERIA
Dr. Bukola Onyekwelu

TOKYO, JAPAN
Prof. Arisa Ema

COMING UP:
NEW DELHI, INDIA
PUEBLA, MEXICO

GLOBAL AI DIALOGUES

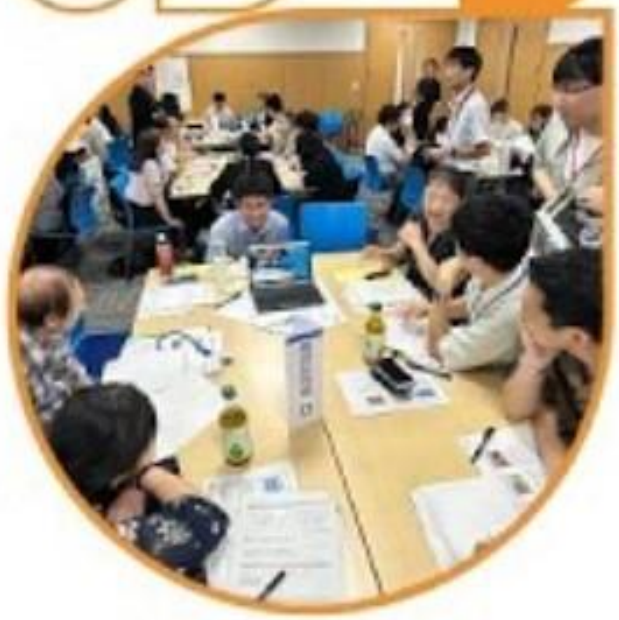




https://www.tc.u-tokyo.ac.jp/en/ai1ec_event/12879/

GLOBAL AI DIALOGUES

TOKYO



Building Multicultural and Multilingual Safe Large Language Models

- Generative AI must reflect cultural and linguistic risks globally.
- Continuous updates needed to identify harmful content specific to each culture.
- "Red teaming" tests AI models by inducing harmful content but is mostly Western-focused.
- Singapore's Infocomm Media Development Authority launched the AI Safety Red Teaming Challenge (Nov 2024).
- Collaboration required among AI researchers, social scientists, policymakers, and practitioners.

https://www.tc.u-tokyo.ac.jp/en/ai1ec_event/13328/

The Collingridge Dilemma

- It is difficult to predict the impact of a technology before it is used in society
- Once a technology has become widespread, it is difficult to control

We may be living in an "experimental society", not a society with a "social experiment”

We need to rethink what kind of society we want to live in