Studies on AI Networking in Japan and A Hypothetical Case (Trolley Problem)

Nov. 17, 2016
The Economic and Social Implications of Artificial Intelligence,
TFF 2016, OECD, Paris
Dean, Grad. Sch. of Pol’y Stud., CHUO UNIV., Tokyo
Prof. Dr. Susuma HIRANO

http://global.chuo-u.ac.jp/english/academics/graduate/policy_studies/

Studies on AI Networking by MIC of Japan (Chronology)

February, 2015
Study Group concerning the Vision of the Future Society Brought by Accelerated Advancement of Intelligence in ICT
June 30, 2015
Report 2015

February, 2016
Conference on Networking among AIs
April 15
Interim Report “Wisdom Network Society (WINS) Produced by the Networking among AIs”
[April 29 and 30
G7 ICT Ministers’ Meeting in Takamatsu, Kagawa]
June 20, 2016
Report 2016 “Impacts and Risks of AI Networking”

Conference toward AI Network Society
Conference on the Networking among AIs

1. Purpose
The Conference deliberated the following items with consideration of the 2040s with regard to the emergence of "AI Networking." (*)
- The image of society to be aimed at and its basic philosophy.
- Evaluation of social and economic impacts and risks caused by AI networking.
- Future challenges (social, economic, and ethical issues).

(*) "AI Networking" indicates an accomplished society; establishment of the AI network system, in which AI is included as a component; advancement of AI network system including coordination among AI network systems.

2. Study Framework
[Chairperson] Osamu Sudo
[Advisor] Jun Murai
[Acting Chairperson] Susumu Hirano

Members Attending the Conference on the Networking among AIs

[Reference] ICT Ministers' Meeting in Takamatsu, Kagawa (April 29 and 30, 2016)

A List of Members Attending the Conference on the Networking among AIs
Idea of "AI Networking" and Stages of Its Progress

The AI Networking

(1) Establishment of "AI Network Systems"(*)
and (ii) advancement thereof thru. the coordination among the AI Network Systems, etc.

(*) The term, "AI Network Systems," means information and communications network systems that include AI as their component.

Stages in progress of the AI Networking

(1) AI functions independently from the Internet, etc. to assist human beings without coordination with other AI. (stand-alone AI)

(2) Networks of AI are formed and autonomous coordination and autonomous harmonization progress in various sectors of society.
   - Various AI with different uses appear on the network.
   - AI with the ability to coordinate multiple AI also appear.
   - Multiple AI collaborate and work in harmony with one another.
   (Examples)
     - Coordination of industrial machinery and construction members, coordination between service robots and sensors.
     - Autonomous coordination of transport, logistics, office work, living environments, etc.

(3) Latent capabilities of human beings are drawn out by the AI network system, resulting in both physical and intellectual development (enhancement of human beings).
   - Coordination of sensors, actuators, AI, and human beings
   - Improved sensory organ capabilities
   - Improved capabilities of the human body
   (Examples) Information in the brain is output externally.
   - Artificial arms, legs, and robots are operated by human thought.
   - Virtually experiencing events in remote locations.
   - Cooperative operation of robots in remote locations.

(4) Coexistence of human beings and the AI network system

"Wisdom Network Society": the Society to Be Built

From "Knowledge" to "Wisdom"

Advanced Information and Telecommunications Network Society (starting 2000)

Wisdom Network Society (WINS)

1. Society to Be Built

- Coexistence of human beings with AI network system
- Free and safe creation, distribution and linking of data, information and knowledge to formulate wisdom networks
- Development of collaboration in various sectors going beyond the individual domains of people, things and activities
- A society capable of more creative and dynamic development

2. Basic Principles of WINS

1. Everyone to enjoy all the benefits of AI network systems.
2. Dignity of humans and autonomy of each individual
3. Innovative research and development and fair competition
4. Controllability and transparency
5. Participation of stakeholders
6. Harmony of physical space and cyberspace
7. Realization of vibrant regional communities through the cooperation beyond space barriers
8. Resolution of global issues under networked cooperation

Wisdom

Ability to use intelligence to deal with issues based on data, information, and knowledge

Knowledge

Systematic accumulation of data and information

Information

Combination of data to provide meaning

Data

Fragmented facts, numerical values, and text

Copyright © 2013 CHUO UNIVERSITY, All Rights Reserved.
Social and Economic Impacts Caused by AI Networking

Social and economic impacts caused by AI networking are evaluated in the time series from 2020s to 2040s in each field of public, life, and industrial areas.

**[Public area]**
- Public infrastructure, disaster prevention, smart cities, public administration.

**[Life area]**
- Life support (personal assistance), creation of richness.

**[Industrial area]**
- Common matters (corporate business, etc.), agriculture, forestry and fisheries, manufacturing, transportation and logistics, wholesale and retail, finance and insurance, medical and nursing care, education and research, service industry, construction.

**[Sample 1] Manufacturing**

It is expected that smart manufacturing processes and supply chains will emerge around 2025, and production optimization and advanced multi-product variable production (customization) will be realized in response to the dynamic balances of demand and supply.

To this end, the use of AI networking systems to monitor abnormal behavior and provide healthcare and independent living assistance will be realized.

**[Sample 2] Medical and nursing care**

The disease prediction of patients based on their vital data and the health management of people based on their genetic information will be realized, thus resulting in an extension of healthy life expectancy.

AI networking will predict the future demand for products from consumers' purchasing behavior.

- Free of manual maintenance.
- Practical achievement of health management using genetic information (2020) [Future]

**Risks Caused by AI Networking (1/2)**

The conference classified social, economic, ethical, and legal risks caused by AI networking as follows*.

1. **Risks Associated with Functions**: Functions that are expected in the AI network system do NOT work appropriately.

2. **Risks related to Legal System, Rights, or Interests**: AI network system infringes rights or interests.

* Some risks have both sides, (example: Risk of accident)

For studying the ideal state of evaluation and management of risks, examination of scenarios that imagine actual applications of AI network system will be needed. (i.e., Risk scenario analysis)

<table>
<thead>
<tr>
<th>Type of Risks</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risks Associated with Functions</strong></td>
<td></td>
</tr>
<tr>
<td>Security-related risks</td>
<td>- Hacking and cyber attack on AI network system.</td>
</tr>
<tr>
<td>- Surrupitious attack on AI network systems without attracting anyone’s attention.</td>
<td></td>
</tr>
<tr>
<td>Risks related to information and communications network systems.</td>
<td>- Occurrence of unintended situation caused by intermingled with various AI in information communication network.</td>
</tr>
<tr>
<td>- Occurrence of unintended situation caused by irregular work of AI when information communication network has some trouble.</td>
<td></td>
</tr>
<tr>
<td>- Data leak and data loss from clouds, and system failure.</td>
<td></td>
</tr>
<tr>
<td>Opacification risks</td>
<td>- As AI algorithm becomes opacified, appropriate control of AI network system becomes difficult for human.</td>
</tr>
<tr>
<td>Risks of control loss</td>
<td>- As AI network system runaways, control by human becomes difficult or impossible.</td>
</tr>
</tbody>
</table>
**Risks Caused by AI Networking (2/2)**

<table>
<thead>
<tr>
<th>Type of Risks</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risks related to Legal Sys., Rights, or Interests</strong></td>
<td></td>
</tr>
<tr>
<td>Risks of accidents</td>
<td>- Accident by the action of an autonomous vehicle or robot on an autonomous decision basis.</td>
</tr>
<tr>
<td>Risks of crimes</td>
<td>- Crime by malware abusing AI network system.</td>
</tr>
<tr>
<td></td>
<td>- Terrorism or crime by autonomous weapon system.</td>
</tr>
<tr>
<td>Risks related to the rights and interests of consumers, etc.</td>
<td>- Inappropriate application of AI network system infringes rights and interests of consumers and young people, etc.</td>
</tr>
<tr>
<td>Risks related to the infringement of privacy and personal information</td>
<td>- As collection and application of personal information by AI network system becomes opacified, control of personal information becomes difficult.</td>
</tr>
<tr>
<td></td>
<td>- AI network systems infringe privacy by surmising people's intentions, health, or future actions, etc.</td>
</tr>
<tr>
<td>Risks related to human dignity and the autonomy of each individual</td>
<td>- AI network systems infringe individual autonomy by invisible manipulation of human's decision making processes.</td>
</tr>
<tr>
<td></td>
<td>- Collapse of the value system of the human-central principles by the technological singularity.</td>
</tr>
<tr>
<td>Risks related to democracy and governance mechanisms</td>
<td>- AI network system's bad influence on voting and people's behavior.</td>
</tr>
<tr>
<td></td>
<td>- As AI network system is applied to the governance of the nation, decision making processes become opacified and the location of responsibility turns ambiguous.</td>
</tr>
</tbody>
</table>

---

**Future Challenges**

1. Formulation of Basic Principles for Research and Development "AI R&D Guidelines"
2. Facilitation of Cooperation toward the Development of AI Networking
3. Securing of Competitive Ecosystem
4. Challenges for Promotion of Economic Development and Innovation
5. Setting Evaluation Indices on Impact of Development of AI Networking and Richness- and Happiness-related Indices
6. Protection of Users
7. Ensuring Security for the AI Networking
8. Institutional Issues relating to Privacy and Personal Data
9. Institutional Issues relating to Content
10. Study of Basic Rules of Society
11. Creating and Sharing Risk Scenarios
12. Accelerated Advancement of Information and Communications Infrastructure
15. Fostering of AI Network System Literacy
16. Personnel Training for the AI Networking
17. Improvements in Working Environments in Response to AI Networking
18. Establishment of a Safety Net
19. Contribution to Human Happiness through the Resolution of Global Issues
20. Approach to Governance of the AI Network Systems
Formulation of AI R&D Guidelines

Similar to the OECD guidelines on the protection of privacy and transborder flows of personal data, the OECD guidelines for the security of information systems and networks, and the OECD guidelines for cryptography policy, it is necessary to initiate discussions and considerations in the OECD and other organizations, with the involvement of the stakeholders concerned, toward formulating international guidelines (tentatively named "AI R&D Guidelines") consisting of principles and their explanations to be taken into account in R&D of AI as a non-binding framework.

【Basic approach to formulation and interpretation of basic principles】
The following approach should be noted in the preamble as the basic approach.

1. A society which could implement the following should be aimed for:
   - The benefits of the AI network system should be received by everyone;
   - Human dignity and personal autonomy should be protected;
   - The controllability and transparency of the AI network system should be ensured; and
   - The AI network system should be used safely and securely.
2. The various risks which might occur in each stage in progress of R&D should be handled timely as well as appropriately.
3. A balance between relevant values and interests should be maintained with the involvement of various stakeholders, with due consideration given to necessity for innovative R&D and fair competition.
4. The guidelines should be reconsidered as needed in response to the progress in the AI Networking and the emergence of risks.

【Content of basic principles】
At least the following should be incorporated into basic principles.

1. Principle of Transparency: Ensuring the abilities to explain and verify the operation of the AI network system should be ensured.
2. Principle of User Assistance: Giving consideration so that the AI network system would assist users and appropriately provide users with opportunities to make choices.
3. Principle of Controllability: Ensuring controllability of the AI network system by humans should be ensured.
4. Principle of Security: Ensuring the robustness and dependability of the AI network system should be ensured.
5. Principle of Safety: Ensuring the robustness and dependability of the AI network system should be ensured.
6. Principle of Privacy: Giving consideration so that the AI network system will not infringe the privacy of users and third parties.
7. Principle of Ethics: Research and development of the AI network system should respect human dignity and personal autonomy.
8. Principle of Accountability: Ensuring accountability of the AI network system should be accountable to related stakeholders such as users by researchers/developers of AI to be networked.

Proposed Principles in "AI R&D Guideline"

1. Principle of Transparency
   - Ensuring the abilities to explain and verify the behaviors of the AI network system
2. Principle of User Assistance
   - Giving consideration so that the AI network system can assist users and appropriately provide users with opportunities to make choices
3. Principle of Controllability
   - Ensuring controllability of the AI network system by humans
4. Principle of Security
   - Ensuring the robustness and dependability of the AI network system
5. Principle of Safety
   - Giving consideration so that the AI network system will not cause danger to the lives/bodies of users or third parties
6. Principle of Privacy
   - Giving consideration so that the AI network system will not infringe the privacy of users or third parties
7. Principle of Ethics
   - Respecting human dignity and individual autonomy in conducting research and development of AI to be networked
8. Principle of Accountability
   - Accomplishing accountability of related stakeholders such as users by researchers/developers of AI to be networked

Reference OECD guidelines governing privacy, security, and so on, it is necessary to begin discussions and considerations toward formulating an international guideline consisting of principles governing R&D of AI to be networked ("AI R&D Guideline") as framework taken into account of in R&D of AI to be networked.

Proposal of Discussion toward Formulation of AI R&D Guideline

Sanee TAKAICHI, Minister for Internal Affairs and Communications, proposed that each G7 country should take the lead under the cooperation of the OECD and other international organizations in working together with stakeholders concerned with the industrial, academic, civil, and governmental sectors to progress international discussions over the socioeconomic impact of AI networking and social, economic, ethical, and legal issues on AI networking, including the formulation of AI development principles*. The participated countries agreed to Minister TAKAICHI’s proposal.

*A tentative plan on AI development principles consisting of eight items was distributed prior to the proposal from Minister TAKAICHI.
Contributions toward Formulating Guidelines

MIC of Japan assembled “the Conference toward AI Network Society” in this October, for the sake of contribution to discussions and considerations in the OECD on social, economic, ethical, and legal issues caused by AI networking. The conference will make a draft of the “AI R&D Guidelines,” for international discussions and considerations in the OECD, as well as analyze details of social and economic impacts and risks caused by AI networking.

In March 2017, in Tokyo, MIC will hold the international symposium on AI networking. In the symposium, Japan intends to introduce the progress in considerations by the Conference toward AI Network Society. Japan would like to hold this symposium to accelerate international communities’ discussions on and formulation of “AI R&D Guidelines.”

Structure and Study Items of the Conference toward AI Network Society

Conference toward AI Network Society

- Issues and institutional matters related to AI research and development principles
- Matters related to the evaluation of impacts and risks caused by AI networking

Subcommittee on AI R&D Principles

- Issues and institutional matters related to AI R&D Principles
- Matters related to AI R&D Principles and guidelines
- Institutional matters related to principles and guidelines

Subcommittee on Impact and Risk Assessment

- Evaluation of positive impacts and challenges in each field
- Evaluation of risks in each field (via Risk Scenario Analysis)
- Other influences caused by AI networking, etc.

Various Kinds of Stakeholders:
(Government, industries, academia, citizens, and international organizations)
The Bridge Problem (1/2)

• Suppose that a school bus suddenly enters into the AV’s lane.

• Suppose that the AV has only two choices;
  (1) to go straight and kill 30-40 children and one driver in the school bus; or
  (2) to turn right and kill the occupant of the AV.

• The manufacturer might prefer (1), while utilitarians might recommend (2).
Application of the AI R&D Principles to the Bridge Problem (1/2)

- **1st Principle of Transparency**: Ensuring the abilities to explain and verify the behaviors of the AI network system
  - Manufacturers might manipulate the AI in a covert manner so that the AV would always choose to protect its occupant(s) by sacrificing the school bus’s 30-40 children. See Noah J. Goodall, Ethical Decision Making during Automated Vehicle Crashes, 2424 TRANSPORTATION RESEARCH RECORD: J. TRANSP. RESEARCH Bd. 58, 63 (2014) (“A self-protection component built into the automated vehicle’s ethics could be hidden in a complicated neural network and discoverable only through the analysis of long-term crash trends. Safeguards must be in place to ensure that such a thing does not happen.” (emphasis added)).

- **2nd Principle of User Assistance**: Giving consideration so that the AI network system could assist users and appropriately provide users with opportunities to make choices
  - The Manufacturer’s choice (1) rather than (2) might be against the occupant’s choice if the manufacturer has done so without taking into account the occupant’s intent. (I.e., Some occupants might prefer the choice (2).)

- **7th Principle of Ethics**: Respecting human dignity and individuals’ autonomy in conducting research and development of AI to be networked
  - Is the choice (1) ethically correct or in compliance with human dignity?

Application of the AI R&D Principles to the Bridge Problem (2/2)

- **8th Principle of Accountability**: Accomplishing accountability to related stakeholders such as users by researchers/developers of AI to be networked
  - Is the choice (1) acceptable to the parents of the 30-40 children?; Could the choice (1) be called a “responsible” act?

Another unsolved issue is:

- **5th Principle of Safety**: Giving consideration so that the AI network system would not cause danger to the lives/bodies of users or third parties
  - The choice (1) would harm third parties’ lives or bodies; The choice (2) would harm users’ lives or bodies.
  - Thus, the 5th Principle, as of now, has not answered the Bridge Problem (which is like the robot in ISAAC ASIMOV’S RUNAROUND (1942) in I, ROBOT (1950)).
http://global.chuo-u.ac.jp/english/academics/graduate/policy_studies/

Thank you for your attention! ;-)