AI Governance – Current Status

2019.4.12

전길남

KAIST

2019 SPRi Spring Conference

2019.3.20rev4.4

Internet Governance, then AI Governance

How do we develop AI Eco-systems?

Table of Contents

1. Internet Governance and Digital Governance

Internet Governance, and Digital Governance

IoT Governance

Data Governance

Cybersecurity Governance

AI Governance

Principles:

Internet Governance Principles

Cybersecurity Norms

AI Principles

2. Is Internet Governance Applicable to AI Governance?

1. **Yes**
   - Open Process with Open Documents
   - Multistakeholders
2. **No**
   - Open Data
   - Standards
   - Government’s Role
3. **Others**
   - Source Code (limited)

3. AI History

1940s Neural Network Model (McCulloch-Pitts), Cybernetics (Wiener)

1950s Turing Test, Dartmouth AI Workshop

1960s Initial AI Boom

1970s AI Winter

1980s 2nd AI Boom; Neural Network, Expert Systems

Late 80s - Early90s 2nd AI Winter

2000s Deep Neural Network/Deep Learning

2010s 3rd AI Boom (Data, Cloud Computing, Algorithm)

4. Narrow AI and General AI

4.1 Narrow AI or Artificial Narrow Intelligence (ANI)

Current Development and Deployment Effort

Deep Learning as Primary Scheme

Pattern Recognition as Primary Area of Applications;

- Games (chess, Go,….)
- Image recognition, Speech recognition, Natural language processing,…

4.2 General AI or Artificial General Intelligence(AGI)

Human Level Intelligence and Super Intelligence

Many Years Away Even If Realized

Overlapping with Brain Science

Nick Bostrom wrote a book on Super Intelligence, and proposed a diagram on general AI and super intelligence.

Stuart Russell and Max Tegmark proposed to prepare for realization of general AI in timely manner.

5. Areas of AI Governance

Social Area:

(1) AI Principles

(2) AI Policy

(3) Social and Economical Impact

(4) Institution

(5) Ethics

Technical Area:

(11) Algorithm

(12) Security

(13) Safety

Long Term Area:

(A) General AI

(B) Existential Risk

(1) AI Principles

We have many good AI principles which are listed below.

Asilomar AI Principles were developed by the participants of Beneficial AI Conference organized by FLI in 2017.

What do we do next in order to have globally accepted AI principles?

List of AI Principles:

1. Asimov's Three Laws of Robotics, 1940.
2. Partnership on AI, Tenets, 2016.
3. Asilomar AI Principles, 2017.
4. Japanese Government, AI R&D Principles, 2017.
5. Montreal Declaration on Responsible AI, 2018.
6. UNESCO, Principles for AI, 2019.3. (to continue at Osaka G20 in 2019.6)

Remark: UN is working on Ban on Lethal Autonomous Weapons through CCW now.

(2) AI Policy

FLI delivered Global Policy Reports in 2018, which covers the following;

- Global AI Policy
- National and International AI Strategies
- AI Policy Challenges and Recommendations
- AI Policy Resources

Medium report by Dutton, An overview of national AI policies in 2018 covered over 20 countries in the world including around 10 Asian countries.

(3) Social and Economical Impacts

AI would have very pervasive impacts to human society and global economy through its deployment in the coming decades.

The impact to the society covers almost every aspect of the human society, and we need to take this matter into consideration.

The economic impact is considered around 15-20% increase on the global economy by 2030s, and most of the benefit would go to the most developed countries and the major AI companies in the world as "winners take all". We need to adjust economical impacts proactively and consider on their distribution.

(4) Institutions

Do we need any global institutions beyond standardization?

Possible Examples in AI include AI Safety?

Examples in other disciplines include IAEA for nuclear technology and IPCC for global warming.

For national cases, we have institutions in some countries now, and more are coming soon.

(5) Ethics

CFI recently wrote a report on AI ethics based on the following developments;

- Asilomar AI Principles (2017)
- ACM Statement on Algorithmic Transparency and Accountability (2017)
- Japan Society on AI, Ethical Guideline (2017)
- Montreal Declaration on Responsible AI Principles (2017-2018)
- IEEE P7000 Committee on Ethical Consideration on AI and AS (2017)
- Partnership on AI Tenets (2018)
- UK House of Lords cross-sector AI Code (2018)
- EC High Level Expert Group: Draft Ethics Guidelines for Trustworthy AI (2018)
- Google's AI Ethics Principles (2018)

Berkman Klein Center/Media Lab created Ethics and Governance of AI Initiative in

(6) Accountability (and Explainability)

FLI Global AI Policy states on Accountability, Transparency and Explainability as follows;

"Holding an AI system or its designers accountable poses several challenges. The lack of transparency and explainability, associated with machine learning in particular, means it can be hard or impossible to know why an algorithm made a particular decision. There is also a question of who has access to key algorithms and how understandable they are, a problem exacerbated by the use of proprietary algorithms. As decision-making is ceded to AI systems, there are not clear guidelines about who would be held accountable for undesirable effects.

DAPRA has Explainable AI Project now.

SNU Annual Conference, AI: Governance and Accountability focuses on accountability.

(11) Algorithm

AI algorithm could bias AI systems, and harm their deployments to human society.

AI algorithm needs to be transparent as much as possible, and they should be explained.

We need to pay good attention on data to be used for AI.

Remark: (5) Accountability and Explainability also covers Algorithm.

(12) Security

Security on AI systems are very important, and AI technology is increasingly used to enhance security. Cybersecurity with AI is very important but raises very difficult issues. Cybersecurity could be enhanced substantially with proper deployment of AI, and we could solve some of the pending issues on the cybersecurity. On the other hand, the same and similar AI technologies could be abused.

Center for Study on AI Governance published articles on AI and cybersecurity including Zwetsloot's AI and International Security.

Center for Long Term Cybersecurity (Cussins), UC Berkeley is working on AI and cybersecurity.

The 2019 National Defense Authorization Act in USA made official a new national Security Commission on AI.

(13) Safety

Asilomar AI Principles on Safety states “AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.”

Future of Life Institute (FLI) has been organizing AI Safety Program for over 30 institutions sponsored by Elon Musk and other organizations since 2015 with emphasis on General AI for the second phase in 2018. FLI covers on AI safety in its Existential Risk webpage.

Victoria Krakovna published articles on AI safety including AI Safety Resources.

Stuart Russell offers the course on AI Safety; Safety and Control for AGI in 2018.

AAAI organized Workshop on AI Safety during AAAI-2019.

AI Governance: A Research Agenda by Allan Dafoe, Center for Governance of AI, gave the overview on AI Safety.

(14) Data

Data are very important to AI development, and AI contributes to generate more data. These factors raise various issues including the following;

- Data governance

  How do we govern data; public data, and data collected by companies?

- Privacy

  How can we ensure privacy against data used by AI systems?

EU's GDPR may be a good starting point on the issue of data vs AI.

(A) General AI

General AI, or Artificial General AI (AGI) needs special attention as many raised issues regarding proper handling of AGI such as Stuart Russell and Max Tegmark.

The main issue is “What to do when AGI exceeds human-level intelligence and reach super intelligence?” “AGI would encounter all of the challenges of narrow AI, but would additionally pose its own risks such as containment” according to FLI’s Global AI Policy. Many consider AGI is possible, and could be realized in this century.

Katja Grace published “When will AI exceed human performance?” in 2017.

Analogy to nuclear technology in the mid-20th century is suggested by Stuart Russell, and recommend to prepare now even though we don’t have good consensus on AGI’s realization.

2019 Beneficial AI Conference focused on General AI.

(B) Existential Risk

Future of Life Institute covers benefits and risks of AI in Existential Risk webpage, and stated in AI Safety of AI Policy Challenges and Recommendations “Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.”

Center for Study on Existential Risks at Cambridge cautioned on existential risks of robots and AI among others in the 21st century.

“The missions of Berkeley Existential Risk Initiative (BERI) is to improve human civilization’s long-term prospects for survival and flourishing. Currently, the main strategy is to take on ethical and legal responsibility, as a grant-maker and collaborator, for projects deemed to be important for reducing existential risk.”[existence.org]

6. Other Issues

1. Analogy: AI Technology to Nuclear Technology – Stuart Russell
2. Developing Countries vs Developed Countries – “Winners Take All”
3. Standards
4. Structuring of AI Governance areas
5. AI Literacy and Capacity Building
6. Autonomous Weapon

References

(1) Articles

(2) Books

(3) Conferences

(4) Courses

(5) Organizations

(6) Presentation Materials

(7) Reports

(8) Video

References - Articles

Kilnam Chon, Digital Governance, APSIG.asia, 2018.

Kilnam Chon, AI: Past and Present, 2018.

Kilnam Chon, AI Governance, 2019.[to be published]

Alan Dafoe, AI Governance: Research Agenda, Center for Study on AI Governance, Oxford, 2018.

Tim Dutton, An overview of national AI strategies, Medium, 2018.6.29.

Edward Felten, AI and Explainability, 2018.

Fortune, (Special Issue on AI including smart speakers), November 2018.

Google, AI Governance – White Paper, 2019.[to be published]

Michael Jordan, AI: revolution has not happened yet, Medium, 2018.4.

Kai-Fu Lee, AI Superpowers, 2018 (in Fortune, November 2018)

FLI, Asilomar AI Principles, 2017.

McKenzie, Promise and Challenge of the Age of AI, 2018.

PwC, Global Artificial Intelligence Study: Sizing the Prize, 2017.

Steven Strogatz, One giant step for a chess-playing machine, NYT, 2018.12.

Max Tegmark, How Far Will AI Go?, 2018.

Wired, How to Teach AI Some Common Sense, 2018.11.

Jess Whittlestone, et al., The role and limits of principles in AI Ethics, AES-19, 2019.

References - Books

Nick Bostrom, Superintelligence, 2014.

Keith Frankish and William Ramsey, Cambridge Handbook of AI, 2014.

Yuval Noah Harari, 21 Lessons for the 21st Century, 2018.

Kai-Fu Lee, AI Superpowers, 2018.

Stuart Russell and Peter Norvig, AI: A modern approach.

Max Tegmark, Life 3.0, 2017.

References - Conferences

ACM/AAAI, AI, Ethics and Society, 2018, 2019. (AAAI Workshop in 2016, 2017)

APSIG, AI Governance Workshop, Annual APSIG Meeting, 2018, 2019.

FLI, Beneficial AI Conference, 2015, 2017, 2019.

Global Government Summit, Global Governance of AI, Dubai, 2019.

IGF, Annual Conferences, 2017, 2018.

KIAS, AI, Ethics and Governance, 2018. (and KAIST with Taming AI in 2018)

Seoul National University, AI: Governance and Accountability, 2017, 2018.

Tokyo University, AI and Society Symposium, 2017.

UN CCW, Group of Gov. Experts on Lethal Weapon Systems, 2019.3, 2019.8.

UNESCO, Principles of AI, Global Conference, 2019.3.

References – Courses

APSIG, AI Governance by Danit Gal.

APSIG, Digital Governance by Kilnam Chon, 2018.

Colombia University, AI, edX.

Coursera, AI for Everybody, 2019.

Microsoft, Introduction to AI, edX.

Microsoft, Ethics and Law, edX.

MIT, Artificial General Intelligence, 6.S099 (by Lex Fridman), 2018.

MIT, Minds and Machines, 24.09 (by Alex Byrne), 2017.

UC Berkeley, Introduction to AI, CS188.

University of Helsinki, Elements of AI

References - Organizations

AI Now Institute, New York University, USA
AI Policy Institute, SNU, South Korea
AI Governance Group, APSIG, Asia
Berkeley Existential Risk Initiative, Berkeley, USA.
Center for Governance of AI, Future of Humanity Institute, Oxford, UK
Center for Human-Compatible Artificial Intelligence, Berkeley, USA
Center for Study on Existential Risks, Cambridge, UK
Ethics and Governance of AI Foundation, USA
Ethics and Governance of AI Initiative, Berkman Klein Center & Media Lab, USA
Future of Life Institute, USA
Future Society, AI Initiative, Harvard University
IEEE, Global Initiative on Ethical Consideration on AI and AS, P7000
ISO/IEC JTC1 SC42, Standardization on AI
Japan Society on Artificial Intelligence, JSAI Ethics Committee
Leverhulme Center for Future of Intelligence, Cambridge (and 3 others), UK

References – Presentation Materials

Kilnam Chon, AI: Past and Present, 2018.

Kilnam Chon, AI Governance, 2019.

Allan Dafoe, AI, Strategy, Policy and Governacne, Beneficial AI, 2019.

Arisa Ema, AI Ethics and Policy, Taming AI, Seoul, 2018.

Danit Gal, AI Governance, APSIG, 2017

Woodraw Herzog, AI: Accountability, 2018.

References - Reports

- AI Index Report. [annual; 2018,…]
- Now Institute, AI Now 2017 Report, 2017.
- EC, Communication from the Commission: AI for Europe, 2018.4.25.
- EU, Guideline for Trustworthy AI, 2018.
- EU, Digital Europe; 2021-2027, 2018.
- Future of Life Institute (FLI), Global AI Policy, 2018.
- Chinese Government, A new generation AI development plan (AIDP), 2017.7.20
- China Standard Administration, White Paper on AI in China, 2018. [English]
- Allan Dafoe, AI Governance: Research Agenda, Center for Governance of AI, FHI, Oxford, 2018.
- Japanese Government (Cabinet Office), AI and Human Society, 2017.
- Japan Society on AI, Ethical Guideline, 2017.
- McKinsey, AI problems and promises, 2018.10. (and more in 2018)

References – Reports (Continued)

- Center for Governance of AI, Oxford, AI: American attitudes and trends, 2019.1.
- PwC's Global Artificial Intelligence Study: Sizing the prize, 2017.
- Roadmap for US Robotics, 2016 Edition; From Internet to Robotics.
- Stanford AI 100 Report, 2015
- Tsinghua University, China AI Development Report, 2018.
- UN CCW Group of Gov. Experts on Lethal Weapon Systems, 2018.
- Web Foundation, Future of technology - AI. [White paper on AI], 2017.
- World Government Summit, Global Governance of AI, 2019.2.

References – Video

Nick Bostrom, Superintelligence, 2014.

APSIG, AI Governance by Danit Gal, 2018.

APSIG, Digital Governance by Kilnam Chon, 2018.

Yuval Harari, 21 lessons for the 21st century, 2018.

Kai-Fu Lee, AI Superpowers, 2018.

FLI, Beneficial AI with Asilomar AI Principles, 2017.

NHK Special, Money World, #2 Work Will Be Gone!, 2018.10.7.

MIT, Artificial General Intelligence, 6.S099 (by Lex Fridman), 2018.

Stuart Russell, Long-term future of AI, MIT AI, 2018.

Max Tegmark, Life 3.0, 2017.

Max Tegmark, How Far Will AI Go?, 2018.

Appendix A: Terminology (from Life 3.0 by Tegmark)

- Narrow Intelligence: Ability to accomplish a narrow set of goals; play chess, drive car
- General Intelligence: Ability to accomplish virtually any goal
- Universal Intelligence: Ability to acquire general intelligence
- General AI (AGI): Ability to accomplish any cognitive task at least as well as humans
- Human-level AI: AGI
- Super Intelligence: General intelligence far beyond human level
- Consciousness: Subjective experience
- Ethics: Principles that govern how we should behave
- Qualia: Individual instances of subjective experience
- Intelligence explosion: Recursive self-improvement rapidly leading to super intelligence
- Singularity: Intelligence explosion

Remark: Strong AI and Weak AI

Appendix B: Narrow AI and General AI, and Weak AI and Strong AI

(From Elements of AI, Helsinki)

Narrow AI refers to AI that handles one task. General AI, or Artificial General Intelligence (AGI) refers to a machine that can handle any intellectual task. All the AI methods we use today fall under narrow AI, with general AI being in the realm of science fiction. In fact, the ideal of AGI has been all but abandoned by the AI researchers because of lack of progress towards it in more than 50 years despite all the effort. In contrast, narrow AI makes progress in leaps and bounds.

A related dichotomy is “strong” and “weak” AI. This boils down to the above philosophical distinction between being intelligent and acting intelligently, which was emphasized by Searle. Strong AI would amount to a “mind” that is genuinely intelligent and self-conscious. Weak AI is what we actually have, namely systems that exhibit intelligent behaviors despite being “mere“ computers.

Appendix C: Governance, National Policy, and Public Trust with Allan Dafoe and Jessica Cussins, FLI, 2018.8.30 by Ariel Conn

“Experts predict that artificial intelligence could become the most transformative innovation in history, eclipsing both the development of agriculture and the industrial revolution. And the technology is developing far faster than the average bureaucracy can keep up with. How can local, national, and international governments prepare for such dramatic changes and help steer AI research and use in a more beneficial direction?”

Backup Pages

What is the Governance of AI?

- **Descriptive definition**: The processes by which decisions are made and implemented. This includes norms, policies, institutions, and laws.

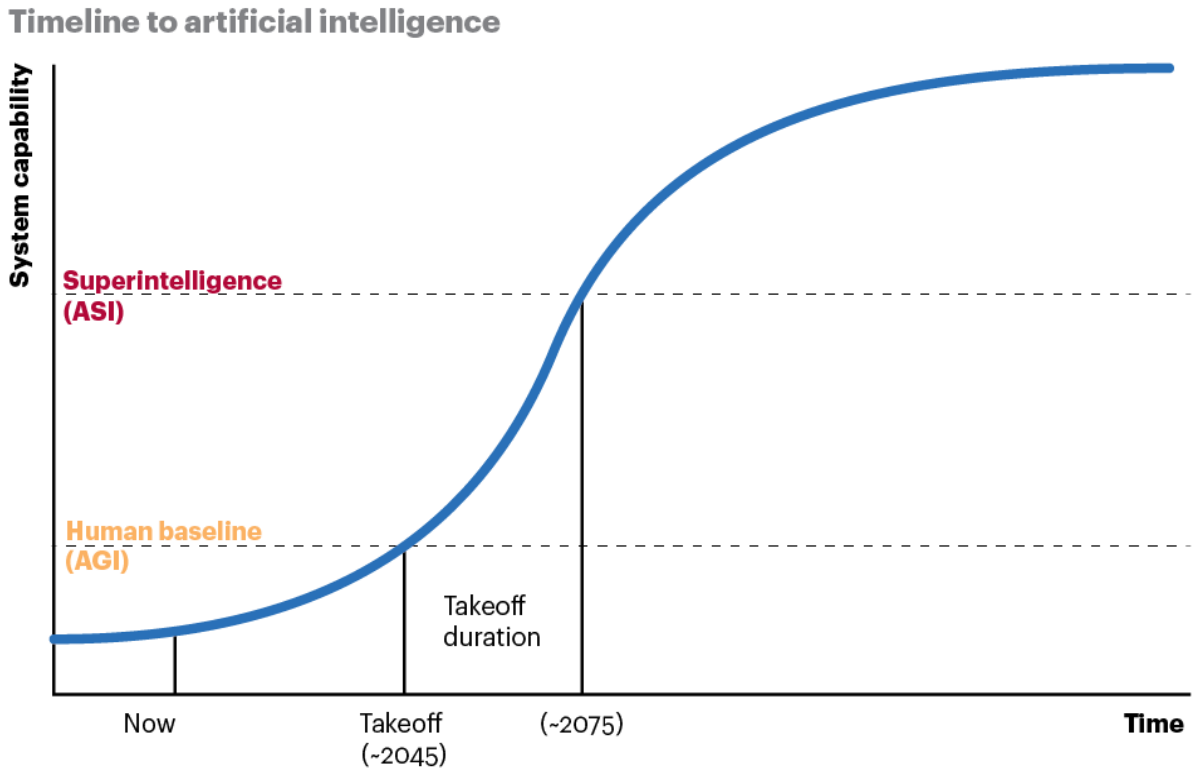
- **Normative definition**: A good set of such processes. Good governance usually means that it is effective, legitimate, inclusive, adaptive.

[From Allan Dafoe, AI Strategy, Policy and Governance, 2019 Beneficial AI]

4.2 AGI (continued): Super Intelligence - Bostrom's Diagram

Figure 19

Developing superintelligent AI may be possible in this century



Note: AI is artificial intelligence, ASI is artificial superintelligence, and AGI is artificial general intelligence.

Sources: WaitButWhy.com, Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*; A.T. Kearney analysis