From Principles to Practices: Building an AI Governance Ecosystem and Framework for Risk Control

Arisa Ema
The University of Tokyo
Japan Deep Learning Association
RIKEN AIP Center
Ministry of Internal Affairs and Communications; The Cabinet Office
“Principles of Human-centric AI society”

A Symposium
AI and SOCIETY
2017.10.10 - 10.11 Wed
Partnership on AI to benefit people and society

http://www.aiandsociety.org/

http://ai-elsi.org/

About AIR
Acceptable Intelligence with Responsibility (AIR) is a research group discussing AI and society. AIR consists of various experts.
We hope to build a society in which people don’t suffer inequality or disadvantage and Information and Communication Technologies (ICT) benefits all people with their diversified values.
Therefore, we aim to create an interdisciplinary network for discussion and encourage research into ICT with the chief goal being to produce prototypes.

Arisa Ema

We aim to construct an interdisciplinary and bottom-up platform that will remain free of the biases created by the profit motive (business) or policy predispositions. Ad-hoc networks involving various stakeholders and experts are created by interesting topics.
What is AI?
Why AI Ethics matters?
Fairness
Accountability
Transparency & Trust
Why it's totally unsurprising that Amazon's recruitment AI was biased against women

Isobel Asher Hamilton  
Oct. 13, 2018, 9:00 AM  
1,468

- Amazon abandoned a project to build an AI recruitment tool, which engineers found was discriminating against female candidates.

- Dr Sandra Wachter, an AI researcher at Oxford
Prediction Fails Differently for Black Defendants

<table>
<thead>
<tr>
<th>Description</th>
<th>White (%)</th>
<th>African American (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled Higher Risk, But Didn’t Re-Offend</td>
<td>23.5%</td>
<td>44.9%</td>
</tr>
<tr>
<td>Labeled Lower Risk, Yet Did Re-Offend</td>
<td>47.7%</td>
<td>28.0%</td>
</tr>
</tbody>
</table>
Adversarial attacks:
All images in the right column are predicted to be an “ostrich, Struthio camelus”

Hackers could easily fool artificial intelligence

"We're entering an era in which our enemies can make anyone say anything at any point in time."
How to govern AI?
1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and Data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability
Regulatory framework proposal on Artificial Intelligence

The Commission is proposing the first ever legal framework on AI, which addresses the risks of AI and positions Europe to play a leading role globally.

The regulatory proposal aims to provide AI developers, deployers and users with clear requirements and obligations regarding specific uses of AI. At the same time, the proposal seeks to reduce administrative and financial burdens for business, in particular small and medium-sized enterprises (SMEs).

The proposal is part of a wider AI package, which also includes the updated Coordinated Plan on AI. Together they guarantee the safety and fundamental rights of people and businesses, while strengthening AI uptake, investment and innovation across the EU.

Why do we need rules on AI?

The proposed AI regulation ensures that Europeans can trust what AI has to offer. While most AI systems pose limited to no risk and can be used to solve many societal challenges, certain AI systems create risks that need to be addressed to avoid undesirable outcomes.

For example, it is often not possible to find out why an AI system has made a decision or prediction and reached a certain outcome. So, it may become difficult to assess whether someone has been unfairly disadvantaged, such as in a hiring decision or in an application for a public benefit scheme.

Although existing legislation provides some protection, it is insufficient to address the specific challenges AI systems may bring.

The proposed rules will:

- address risks specifically created by AI applications
- propose a list of high-risk applications
- set clear requirements for AI systems for high-risk applications
- define specific obligations for AI users and providers of high-risk applications
- propose a conformity assessment before the AI system is put into service or placed on the market
- propose enforcement after such an AI system is placed in the market
- propose a governance structure at European and national level

A risk-based approach
What are the OECD Principles on AI?

The OECD Principles on Artificial Intelligence promote artificial intelligence (AI) that is innovative and trustworthy and that respects human rights and democratic values. They were adopted on 22 May 2019 by OECD member countries when they approved the OECD Council Recommendation on Artificial Intelligence. The OECD AI Principles are the first such principles signed up to by governments. Beyond OECD members, other countries including Argentina, Brazil, Colombia, Costa Rica, Peru and Romania have already adhered to the AI Principles, with further adherents welcomed.

The OECD AI Principles set standards for AI that are practical and flexible enough to stand the test of time in a rapidly evolving field. They complement existing OECD standards in areas such as privacy, digital security risk management and responsible business conduct.

The Japanese Cabinet Office: Social Principles of Human-centric AI (Draft)

• Fundamental Philosophy
  • (1) Dignity, (2) Diversity & Inclusion, (3) Sustainability

• Social Principles of Human-Centric AI
  1. Human-centric
  2. Education/Literacy
  3. Privacy Protection
  4. Ensuring Security
  5. Fair Competition
  6. Fairness, Accountability, and Transparency
  7. Innovation

https://www.cas.go.jp/jp/seisaku/jinkouchinou/
R& D Principles
• Do Good
• For Humanity
• Be Responsible
• Control Risks
• Be ethical
  • Making the system as fair as possible, reducing possible discrimination and biases ...
• Be Diverse and Inclusive
• Open and Share

Use
• Use wisely and properly
• Informed-consent
• Education

Government
• Optimizing Employment
• Harmony and cooperation
• Adaption and moderation
• Subdivision and implementation
• Long-term planning

https://www.baai.ac.cn/blog/beijing-ai-principles
IEEE Ethically Aligned Design 1st edition

1. From Principles to Practice
2. General principles
3. Classical Ethics in AI/IS
4. Well-being
5. Affective Computing
6. Personal Data and Individual Agency
7. Methods to Guide Ethical Research and Design
8. A/IS for Sustainable Development
9. Embedding Values into Autonomous and Intelligent Systems
10. Policy
11. Law
Guidelines for Japanese Society for Artificial Intelligence researchers (2017)

1. Contributions to humans
2. Abidance of laws and regulations
3. Respect for the privacy of others
4. Fairness
5. Security
6. Act with integrity
7. Accountability
8. Communication with society and self-development
9. Abidance of ethics guidelines by AI
   • AI must abide by the policies described above in the same manner as the members of JSAL, in order to become a member or a quasi-member of society.
# Partnership on AI: Thematic Pillars

<table>
<thead>
<tr>
<th></th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Safety-Critical AI</td>
</tr>
<tr>
<td>2</td>
<td>Fair, Transparent, &amp; Accountable AI</td>
</tr>
<tr>
<td>3</td>
<td>AI, Labor &amp; the Economy</td>
</tr>
<tr>
<td>4</td>
<td>Collaborations between People &amp; AI Systems</td>
</tr>
<tr>
<td>5</td>
<td>Social &amp; Societal Influences of AI</td>
</tr>
<tr>
<td>6</td>
<td>AI &amp; Social Good</td>
</tr>
</tbody>
</table>

[https://www.partnershiponai.org/](https://www.partnershiponai.org/)
Global companies’ AI principles

• Global Company
  • Sony Group AI Ethics Guidelines (2018.9)
  • Fujitsu Group AI Commitment (2019.2)
  • NEC Group AI and Human Rights Principles (2019.4)
  • NTT Data AI Guidelines (2019.5)

• Start-ups
  • ABEJA launched 3rd party committee “Ethical approach to AI” (2019.7)

AI governance is competitive and collaborative area: Companies need to have AI guidelines and governance frameworks for the management strategy, not CSR
• The case of Dr. Gebru at Google in the U.S. has ironically become that AI ethics are used merely as a **smokescreen to hide their transgressions**

https://ifi.u-tokyo.ac.jp/en/project-news/9008/
Principles

- How
- What
- Who

Practices

- Best Practices
- Governance Ecosystem
- Technologies
<table>
<thead>
<tr>
<th>Chinese Academy of Sciences</th>
<th>Harvard University's Berkman Klein Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humanity</td>
<td>International Human Rights</td>
</tr>
<tr>
<td>Collaboration</td>
<td>Promotion of Human Values</td>
</tr>
<tr>
<td>Share</td>
<td>Professional Responsibility</td>
</tr>
<tr>
<td>Fairness</td>
<td>Human Control of Technology</td>
</tr>
<tr>
<td>Transparency</td>
<td>Fairness and Non-discrimination</td>
</tr>
<tr>
<td>Privacy</td>
<td>Transparency and Explainability</td>
</tr>
<tr>
<td>Security</td>
<td>Safety and Security</td>
</tr>
<tr>
<td>Safety</td>
<td>Accountability</td>
</tr>
<tr>
<td>Accountability</td>
<td>Privacy</td>
</tr>
<tr>
<td>AGI</td>
<td></td>
</tr>
</tbody>
</table>

https://arxiv.org/abs/1812.04814

http://wilkins.law.harvard.edu/misc/PrincipledAI_FinalGraphic.jpg
Principles

Practices

Best Practices

- Utilizations
  - Medicine
  - Finance
  - Risk Prevention
  - Agriculture
  - Military
- Work /働き方
  - Human-machine interaction
  - Future work

Governance Ecosystems

- Companies
- Private Sector and Industry Associations
- Academia
- Public Sector
- User

Technologies

- Trustworthy AI
  - Fair algorithm and Data
  - Explainable AI
  - Sustainable AI
  - Robust AI
- Next AI & Environment
  - 5G Network
  - Neuroscience & Quantum
Contract Guidelines on Utilization of AI and Data Version 1.1 Formulated

December 9, 2019

The Ministry of Economy, Trade and Industry (METI) has released an updated version of the contract guidelines on the utilization of AI and data. This update reflects the revision of related laws and regulations, and the contract guidelines section of the guidelines was formulated in June 2018. The updated version includes revised sections and additional information.

Outline of Data Section

In this guidance, data contract is divided into three types. Each type is analyzed with three aspects such as Structure, main legal issues in a contract, appropriate ways to make a contract.

1. “Providing Data Type”
   - Structure: Providing data from the data provider to the other party (in the case where only the data provider holds the target data)
   - Main Legal Issues in a Contract: Responsibility that the provided data is not what was expected.
   - Appropriate ways to decide: Checklist of contents in the guidance. Sample clauses in the guidance.

2. “Creating Data Type”
   - Structure: Where data is newly created due to involvement of multiple parties, a contract for negotiating the use authority of the data among the parties involved in data creation.
   - Main Legal Issues in a Contract: Terms of usage set up between parties.
   - Appropriate ways to decide: Checklist of contents in the guidance.

3. “Sharing Data Type”
   - Structure: Terms of contracts aimed at sharing data through the platform.
   - Appropriate ways to decide: Major items shown in the guidance.

AI governance Ecosystem

Companies
- Corporate vision
- AI Principles
- Risk Assessment
- Risk Control
- Employee/ Employer Education

Private Sector and Industry Association
- Guidelines
- Standards
- Audit
- Insurance
- Fact Check

Public Sector
- Hard Law, Soft Law
- Third-Party Incident Committee
- Whistleblower System /

Academia

Users
- Education
  - Engineer
  - Experts
  - Policy makers
  - General publics
## AI governance and its evaluation Study Group

<table>
<thead>
<tr>
<th>No.</th>
<th>Topic</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Trend of AI Governance</td>
<td>7/31</td>
</tr>
<tr>
<td>2</td>
<td>Ethics Guidelines</td>
<td>8/25</td>
</tr>
<tr>
<td>3</td>
<td>Audits and assurance</td>
<td>9/25</td>
</tr>
<tr>
<td>4</td>
<td>Insurance</td>
<td>10/26</td>
</tr>
<tr>
<td>5</td>
<td>Standardization</td>
<td>11/24</td>
</tr>
<tr>
<td>6</td>
<td>Internal whistleblowing system / Consumer protection</td>
<td>12/9</td>
</tr>
<tr>
<td>7</td>
<td>Accidents investigation / Third party committee</td>
<td>12/10</td>
</tr>
<tr>
<td>8</td>
<td>Policy and social system 1: Data utilization</td>
<td>12/15</td>
</tr>
<tr>
<td>9</td>
<td>Policy and social system 2: Governance guidelines</td>
<td>2/18</td>
</tr>
<tr>
<td>10</td>
<td>Monitoring / Advertising</td>
<td>2/25</td>
</tr>
<tr>
<td>11</td>
<td>Open Data</td>
<td>4/13</td>
</tr>
<tr>
<td>12</td>
<td>Security</td>
<td>4/22</td>
</tr>
<tr>
<td>13</td>
<td>Practices of Companies 1: Traditional companies</td>
<td>4/27</td>
</tr>
<tr>
<td>14</td>
<td>Practices of Companies 2: Start-ups</td>
<td>5/11</td>
</tr>
</tbody>
</table>

[Visit the website](https://www.jdla.org/en/en-about/en-studygroup/en-sg01/)
Risk Chain Model to consider AI governance

Companies

Corporate vision and values

Consider AI Service Requirements & Technologies

Create Risk Scenarios

Draw Risk Chains

Consider Risk Control

- Usage Policy
- Education
- Data modification

AI Principles
- Facial Recognition
- Building Entrance
- Data bias & wrong feedback
- Influence of the noise
Risk Scenarios: Fairness

Trustworthiness: Risk Factors
- Quality
- Ethics
- Usability

Possibility of Risk Realization
- Life
- Right/Property
- Mental/Honor

Service Provider

AI System
- Data
  - Data Quality
  - Data Balance
- AI Model
  - Accuracy
  - Robustness
  - Interpretability
- Application
  - Process Integrity
  - Connectivity
- System Environment
  - Capability
  - Stability
  - Confidentiality
  - Availability
- Communication
  - Consensus
- Traceability

Users

Understand
- User Responsibility
- Autonomy
- Understanding

Expectation
- Expectation

Action
- Self-Defense

Option
- Proper Use

User Environment
- User Ability
- Awareness

Limitation
- Option

Operation
- Scalability
- Sustainability
- Agility
- Safety
- Accessibility
- Auditability

Code of Conduct
- Accountability
- Dignity
- Privacy
- Transparency

Case examples

• Case study
  • Case01. Recruitment AI (2021/07)
  • Case02. Unstaffed convenience stores (2021/09)
  • Case03. Power Line Inspection AI (2021/09)
  • Case04. Defect Detection AI (2021/09)
  • Case05. Road Guide Robot (2021/09)
  • Case06. Verification of Recidivism Possibility AI (2021/09)

• Coming soon in February
  • Cancer detection AI
  • Loan screening AI
  • Smart Home application
  • Driverless bus
  • Plant operation guide

Risk control training using RCM

The process of creating the RCMModel with stakeholders, both those involved in AI practices and those who are not, was itself a learning and awareness experience for both parties.

Through the discussion process, we reaffirmed the values and objectives we wanted to achieve.

It was good that the thought processes we discussed are structured and stored as a visual record.
Thank you for your attention

Arisa Ema
https://arisaema0.wordpress.com/