

Jan. 7th 2020 @ 内閣府

人間中心のAI社会原則とAI政策

須藤 修

* 東京大学大学院教授

* 東京大学総合教育研究センター長

Osamu Sudoh, Prof. and PhD.,
The Univ. of Tokyo

Osamu Sudoh 2020

人間中心のAI社会原則

大阪で開催されたG20において承認された(2019年6月28日)

須藤 修 (東京大学)

内閣府「人間中心のAI社会原則会議」議長

人間中心のAI社会原則（人間中心、教育・リテラシー）

(1) 人間中心の原則

- AIは、人間の労働の一部を代替するのみならず、高度な道具として人間を補助することにより、人間の能力や創造性を拡大することができる。
- AIの利用にあたっては、人が自らどのように利用するかの判断と決定を行うことが求められる。AIの利用がもたらす結果については、問題の特性に応じて、AIの開発・提供・利用に関わった種々のステークホルダーが適切に分担して責任を負うべきである。

(2) 教育・リテラシーの原則

- AIの利用者側は、その概要を理解し、正しく利用できる素養を身につけていることが望まれる。特に、AIに関わる政策決定者や経営者は、AIを社会的に正しい利用ができる知識と倫理を持っていなければならない。
- AIについての教育・リテラシーを育む教育環境が全ての人に平等に提供されなければならない。
- 我々の社会においては誰でもAI、数理、データサイエンスの素養を身につけられる教育システムとなっているべきであり、全ての人が文理の境界を超えて学ぶ必要がある。

人間中心のAI社会原則（プライバシー、セキュリティ）

(3) プライバシー確保の原則

- パーソナルデータを利用したAI及びそのAIを活用したサービス・ソリューションにおいては、政府における利用を含め、個人の自由、尊厳、平等が侵害されないようにすべきである。
- AIの使用が個人に害を及ぼすリスクを高める可能性がある場合には、そのような状況に対処するための技術的仕組みや非技術的枠組みを整備すべきである。

(4) セキュリティ確保の原則

- 少なくとも現在想定できる技術の範囲では、意図的な攻撃に対してAIが常に適切に対応することは不可能であり、セキュリティに対する新たなリスクも生じる。社会は、全体として社会の安全性及び持続可能性が向上するように務めなければならない。
- 社会は、AIの利用におけるリスクの正しい評価やそのリスクを低減するための研究等、リスク管理のための取組を進めなければならない。

人間中心のAI社会原則(公正競争、FAT)

(5) 公正競争確保の原則

- 特定の国にAIに関する資源が集中した場合においても、その支配的な地位を利用した不当なデータの収集や主権の侵害が行われる社会であってはならない。
- AIの利用によって、富や社会に対する影響力が一部のステークホルダーに不当過剰に偏る社会であってはならない。

(6) 公平性、説明責任及び透明性の原則

- AIの設計思想の下において、人々がその人種、性別、国籍、年齢、政治的信念、宗教等の多様なバックグラウンドを理由に不当な差別をされることなく、全ての人々が公平に扱われなければならない。
- AIを利用しているという事実、AIに利用されるデータの取得方法や使用方法、AIの動作結果の適切性を担保する仕組みなど、用途や状況に応じた適切な説明が得られなければならない。
- 人々がAIの提案を理解して判断するために、AIの利用・採用・運用について、必要に応じて開かれた対話の場が適切に持たれなければならない。

人間中心のAI社会原則(イノベーション)

(7) イノベーションの原則

- Society 5.0を実現し、AIの発展によって、人も共に進化していくような継続的なイノベーションを目指すため、国境や産学官民、人種、性別、国籍、年齢、政治的信念、宗教等の垣根を越えて、幅広い知識、視点、発想等に基づき、人材・研究の両面から、国際化・多様化と産学官民連携を推進すべきである。
- AIを効率的かつ安心して社会実装するため、AIに係る品質や信頼性の確認に係る手法、AIで活用されるデータの効率的な収集・整備手法、AIの開発・テスト・運用の方法論等のAI工学を確立しなければならない。
- AI技術の健全な発展のため、プライバシーやセキュリティの確保を前提としつつ、あらゆる分野のデータが独占されることなく、国境を越えて有効利用できる環境が整備される必要がある。
- 政府は、AI技術の社会実装を促進するため、あらゆる分野で阻害要因となっている規制の改革等を進めなければならない。

OECD AI Principles

OECD理事会勧告2019年5月22日調印

PRINCIPLES FOR RESPONSIBLE STEWARDSHIP OF TRUSTWORTHY AI

1.1. Inclusive and sustainable growth and well-being

1.2. **Human-centred values and fairness**

1.3. Transparency and explainability

1.4. Robustness and safety

1.5. Accountability

NATIONAL POLICIES FOR TRUSTWORTHY AI

2.1. Investing in responsible AI research and development

2.2. Fostering an enabling digital ecosystem for AI

2.3. Providing an agile policy environment for AI

2.4. **Building human capacity and preparing for job transformation**

INTERNATIONAL COOPERATION FOR TRUSTWORTHY AI

Skills to use AI

Education policy is expected to require adjustments to expand lifelong learning, training and skills development. As with other areas of technology, AI is expected to generate demand in three skills areas.

- First, **specialist skills** will be needed to program and develop AI applications. These could include skills for AI-related fundamental research, engineering and applications, as well as data science and computational thinking.
 - Second, **generic skills** will be needed to leverage AI, including through AI-human teams on the factory floor and quality control.
 - Third, AI will need **complementarity skills**. These could include leveraging human skills such as critical thinking; creativity, innovation and entrepreneurship; and empathy.
- OECD ed. [2019] Artificial Intelligence in Society, p.110.

European Commission and other European institutions, 2018—

- 「設計による倫理」(Ethics by Design)と規制の枠組み: 倫理的枠組みは、**High-Level Expert Group on Artificial Intelligence (AI HLEG)**によって開発されたAI倫理ガイドラインに基づいている。ECは、加盟諸国に対して調達政策を通じて「設計による倫理」を定着させることを求めている。
- スキルと生涯学習: 才能、スキル、生涯学習を促進する。
- データ: 公益のデータ、健康データを含むAIの産業データのためのプラットフォームへのアクセスを促進するために、Common European Data Spaceの創設を提言している。

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS

Building Trust in Human-Centric Artificial Intelligence

April 8th 2019

The guidelines postulate that in order to achieve ‘trustworthy AI’, three components are necessary: (1) it should comply with the law, (2) it should fulfil ethical principles and (3) it should be robust.

The guidelines identify seven key requirements that AI applications should respect to be considered trustworthy.

- ✓ Human agency and oversight
- ✓ Technical robustness and safety
- ✓ Privacy and data governance
- ✓ Transparency
- ✓ Diversity, non-discrimination and fairness
- ✓ Societal and environmental well-being
- ✓ Accountability

The MIT Schwarzman College of Computing Open in September 2019



The MIT Schwarzman College of Computing was announced last October 2018 as a \$1 billion commitment to addressing the opportunities and challenges presented by the prevalence of computing and the rise of artificial intelligence. **The initiative aims to reorient MIT to bring the power of computing and artificial intelligence to all fields of study, and to educate students in every discipline to responsibly use and develop these technologies.** The college is slated to open in September 2019.

World Economic Forum launches Global AI Council



The World Economic Forum established the “Artificial Intelligence Council”, which has been co-chaired by Sinovation Ventures CEO **Kai-Fu Lee** and Microsoft President **Bradford L. Smith** (May 2019).

“Artificial Intelligence brings great benefits. But people are now facing new challenges, such as ethical issues with AI, personal data protection, and replacement of human job posts. The AI Council is committed to helping people solve this problem,” said Lee during his speech.

✓ April 2020, WEF Conference for AI Strategy will be held at San Francisco.

Some of Portfolio Projects, WEF

- ✓ Unlocking Public Sector Procurement
- ✓ Generation AI
- ✓ Teaching AI Ethics and Social Impact
- ✓ Unlocking Data for Democratizing AI Innovation
esp. Building a data-sharing ecosystem
- ✓ And so on

UNESCO High Level Meeting on Principle for AI 4 March, 2019 UNESCO HQ, Paris



UNESCO事務局長： オードレ・アズレー



Session on Towards a Human-Centred Ethical AI?

- ✓ Osamu Sudoh talked about!
- AI should be developed and deployed **to expand the capabilities of human and to pursue the diverse happiness of various people all over the world.**
- In the AI-based society, we will deploy literacy education on using AI, so as not to over-depend on AI or not to make human decisions wrong by abusing AI.
- AI can expand human capability and creativity not only by replacing part of human task but also by assisting human as an advanced instrument.



Session on Towards a Human-Centred Ethical AI?

- ✓ Osamu Sudoh talked about!
- In order to build human talent of creativity, the educational systems enable everyone **to learn AI algorithm, and so-called STEM education as well as law studies, ethics around AI.**
- And also I am sure that the concept of human-centric is supported by human creativity and compassion, as Dr. Kai-Fu Lee, he is very famous Chinese founder, he has suggested significance of human compassion strongly.



TICAD2019 AIシンポ DFFTの意義

United Nations Educational, Scientific and Cultural Organization
TICAD7 公式サイドイベント
AIの活用に関するパネルディスカッション
Side Event for TICAD7: Panel Discussion on the application of AI
2019年8月29日(木) 10:30~12:00 パシフィコ横浜内 ANNEXホール F204
29th August, 2019 10:30~12:00 F204 ANNEX Hall, Pacifico Yokohama

Speakers

須藤 雅彦
東京大学大学院情報学環センター長
東京大学大学院情報学環教授(経済学)

高橋 博之
国立情報科学研究所所長
東京大学大学院情報学環教授(工学)

佐藤 光正
国立情報科学研究所副所長
東京大学大学院情報学環教授(工学)

セリョウ・ワグネル・ゴードン
ユネスコ・アフリカ地域開発センター
科学文化部長

ドローシー・ゴードン
ユネスコ・アフリカ地域開発センター
科学文化部長

アヌ・バング
ユネスコ・アフリカ地域開発センター
科学文化部長

人工知能は科学技術、特にデータ集約の研究分野で重要な役割を果たしています。しかし、AI技術の活用による具体的な成果は、私たちのニーズを踏まえて実用的な分野に適用されることによって達成されるべきです。今回のTICAD7 AIシンポジウムでは、アフリカからAIの発展と課題、ユネスコのAIにおける役割、そして日本の経験を取り入れたAIの活用促進に向けた協力を取り上げ、特に防災・減災に関連する議論を行います。

AI is already playing an important role in science and technology research, especially in data-intensive fields. However, tangible outcomes of utilizing AI technologies will depend on how we will apply them to practical areas of our needs. This TICAD7 panel discussion will highlight African AI prospects and challenges, UNESCO's role in AI, and Japan's multiple efforts and potential role in advancing the use of AI, with a special focus on disaster risk reduction (DRR).

入場料 無料 定員 80人(先着順・申込不要)
Language: English and Japanese (simultaneous interpretation)
Admission: free Capacity: 80 (No registration required)

Access 〒220-0012 横浜市西区みなとみらい1-1-1
1-1 Minato Mirai, Nishi-ku, Yokohama 220-0012, Japan

会場: ANNEXホール F204
交通: 京浜東北線 新横浜駅 徒歩10分
京浜東北線 横浜駅 徒歩15分
京浜東北線 磯子駅 徒歩15分
京浜東北線 磯子駅 徒歩15分

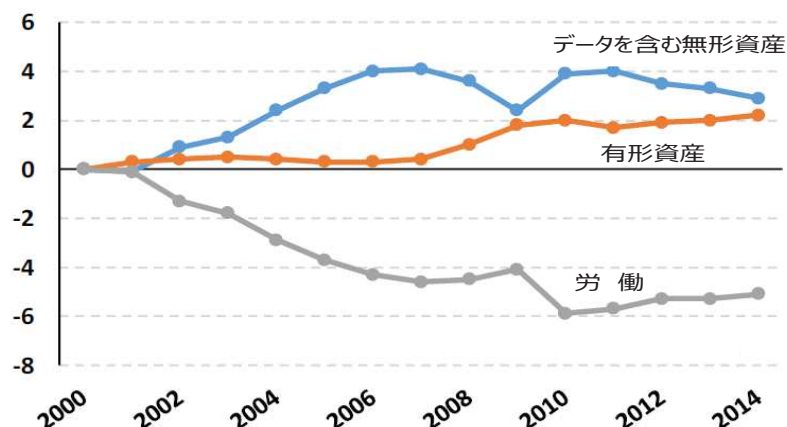
主催: 国連教育科学文化機関 (UNESCO) / 外務省
Organized by: United Nations Educational, Scientific and Cultural Organization (UNESCO) / Ministry of Foreign Affairs of Japan

お問い合わせ: Information E-mail: ticad7sideevent@unESCO.org.jp

- 2019年8月28日ー30日、横浜で第7回TICAD(アフリカ開発会議)が開催された。
- その一環として8月29日、Panel Discussion on the application of AIが開催された(須藤はモデレータ)。
- ドロシー・ゴードンをはじめアフリカやユネスコの代表は、データフローの在り方を変革しなければならないことを強調した。日本の主張しているDFFTが高く評価された。今後、DFFTの詳細を検討し、実行に移さなければならない。

世界の富の分配について : 新たなMetricsが必要!

無形資産・有形資産・労働の分配率の変化幅
先進国の労働分配率: 約51%(IMF、2017年)
発展途上国の労働分配率: 約39.5%(IMF、2017年)



出所: Chen, Los & Timmer (2018) Factor Incomes in Global Value Chains: The Role of Intangibles, Working Paper 25242.

China (2017), *Guideline on Next Generation AI Development Plan*, Government of China, State Council

- 2017年、中国の国務院は、次世代AI開発計画に関するガイドラインをリリースした。AIについて、中国の長期的な視点と各期間の産業目標を提示している。
 - ✓ 2030年までに中国がグローバルなAIイノベーションセンターになる。
- 第13次5カ年計画(2016-20)で、中国は「AI 2.0」を含むScience and Technology Innovation 2030 Megaprojectsという方針を持っている。この計画において、AIベースのビジョン、音声および生体認証、ヒューマンマシンインターフェイスなど、AIハードウェアおよびソフトウェアの研究開発を加速するよう中国企業に求めている。
- 2019年、次世代AI戦略諮問委員会の下に、「次世代AIガバナンス特別委員会」の設置を決定した。この特別委員会において、AIに関する法律、倫理、標準化、社会問題の研究を行う。

Gary Liu is CEO of the *South China Morning Post*, a leading news media company that has reported on China and Asia

- ✓ 「中国のインターネットを見れば、その広範な検閲とディストピア的監視の懸念に気付かずにはいられない。例えば中国は全国民を対象とした社会的信用格付けを展開しようとしており、誠実さや正直さといった、質的な特性に基づいて国民に報酬を与えたり制限を課したりしようとしている。また中国は1億7千万台ある監視カメラの多くに顔認識システムを導入しつつある。」

- ✓ 「それにもかかわらずインターネットは発達し続けている。＜略＞ 2017年末までに中国のインターネット利用者数は7億7200万人に達した。これは 米国、ロシア、ドイツ、イギリス、フランス、カナダの人口を合わせたよりも多い。そのうち98%がスマホ利用で、92%がメッセージングアプリを使用している。いまやネットニュースの利用者が6億5000万人、動画サービス視聴者が5億8000万人もいる。」



Chinese Social Credit Score

- The Chinese social credit score can influence decisions like the deposit level on an apartment rental or online dating matches. A person playing video games for hours every day might, for example, obtain a lower social credit score than a person purchasing diapers who is assumed to be a responsible parent.
 - A broader Chinese social credit system to score the “trustworthiness” of individuals, businesses and government officials is planned to be in place by 2020.
- OECD ed. [2019] Artificial Intelligence in Society, p.56.