

拡張される社会： 人とAIの協力のデザイン



西片健郎

CONTENTS

- I 背景と論点
- II 社会システムの再考を迫るテクノロジーの進化と環境の変化
- III 現在の「機械的な社会システム」が抱える問題
- IV 「機械と生物から学ぶ社会システム」の提案
- VI 日本への示唆とステークホルダーに求められる視点

要約

- 1 サイバー空間の変容とAIの急速な進歩により、社会システムの再考が迫られている。
- 2 現在の社会システムはいわば「機械的な」構造を有している。それは経済価値の創出には大きく貢献するが、社会価値との両立が困難になるという構造的な課題を抱えている。テクノロジーを社会システムの能力を拡張する道具と捉えた場合、「機械的な社会システム」の拡張にAIを中心とする強力なテクノロジーを利用することは、かえって課題を深刻化させるとともに、AI中心の社会をつくり出してしまうリスクがある。
- 3 このような課題を克服する新たな社会システムとして、「生物的な」概念を取り入れた「機械と生物から学ぶ社会システム」を提案する。機械と生物から学ぶ社会システムでは、経済価値と社会価値の両立と価値向上が期待できる。また、「生物的な社会システム」の拡張にAIを中心とするテクノロジーを活用することで、人間と環境中心の社会の実現が期待できる。
- 4 世界が「機械と生物から学ぶ社会システム」へ転換していくうえで、文化的に「生物的な社会システム」との親和性が高い日本は、この領域において世界をリードできるポテンシャルがある。その実現のためには、ある意味で、文化を担う市民が新たな社会システムを自発的につくり出せるような仕組みづくりが求められる。

I 背景と論点

1 背景

筆者は、エンジニアとしてのバックグラウンドを基に、近年は「デジタル社会インフラの技術アーキテクチャ」を軸として、情報や金の流れを規定するさまざまな技術（インターネット、データ、ブロックチェーン、デジタル通貨、ソーシャルメディアなど）の調査や研究開発に携わってきた。先端技術の研究開発の場や国際標準化の場に技術をつくる当事者として参加することを通じて、そのダイナミズムを肌で感じてきた。さらに、社会や経済の専門家と協働することで、そのあるべき姿を模索してきた。

こうした活動を通じて、近年、疑問に思っていたことがある。それは、技術進歩はますます加速し、社会システムの根幹を揺るがすようになっている中、既存の社会システムが技術の変化に追いつけているかということである。現在の社会システムの枠組みの中で、AIを中心とするこれまで以上に強力なテクノロジーをわれわれは扱うことができるだろうか。AIの便益とリスク、活用と規制などの各論や技術論に入る前に、前提となる社会システムのあり方自体について考察することこそが、喫緊の課題ではないだろうか。

2 本稿で示す3つの論点

このような問題意識を踏まえて、本稿では大きく3つの点を論じる。

第1に、サイバー空間の変容とAIの急速な進歩という2つの環境変化により、社会システム^{注1文献1}の再考が迫られていることを示す。第2に、現在の社会システムは課題を抱

えており、テクノロジーを社会システムの能力を拡張する道具と捉えた場合^{注2}、現在の社会システムの拡張にAIを利用することは、課題を増幅することを示す。第3に、課題を克服するための新たな社会システムの姿を提案し、その社会システムの拡張にAIを用いることを提案する。

なお、本稿で新たな社会システムを提案する目的は、未来予測をすることでもベストな社会システムを主張することでもない。AIを中心とする強力なテクノロジーが進歩する未来に向けて、どのような社会システムをつくるべきかという問題を提起するとともに、議論を喚起することにある。

II 社会システムの再考を迫る テクノロジーの進化と 環境の変化

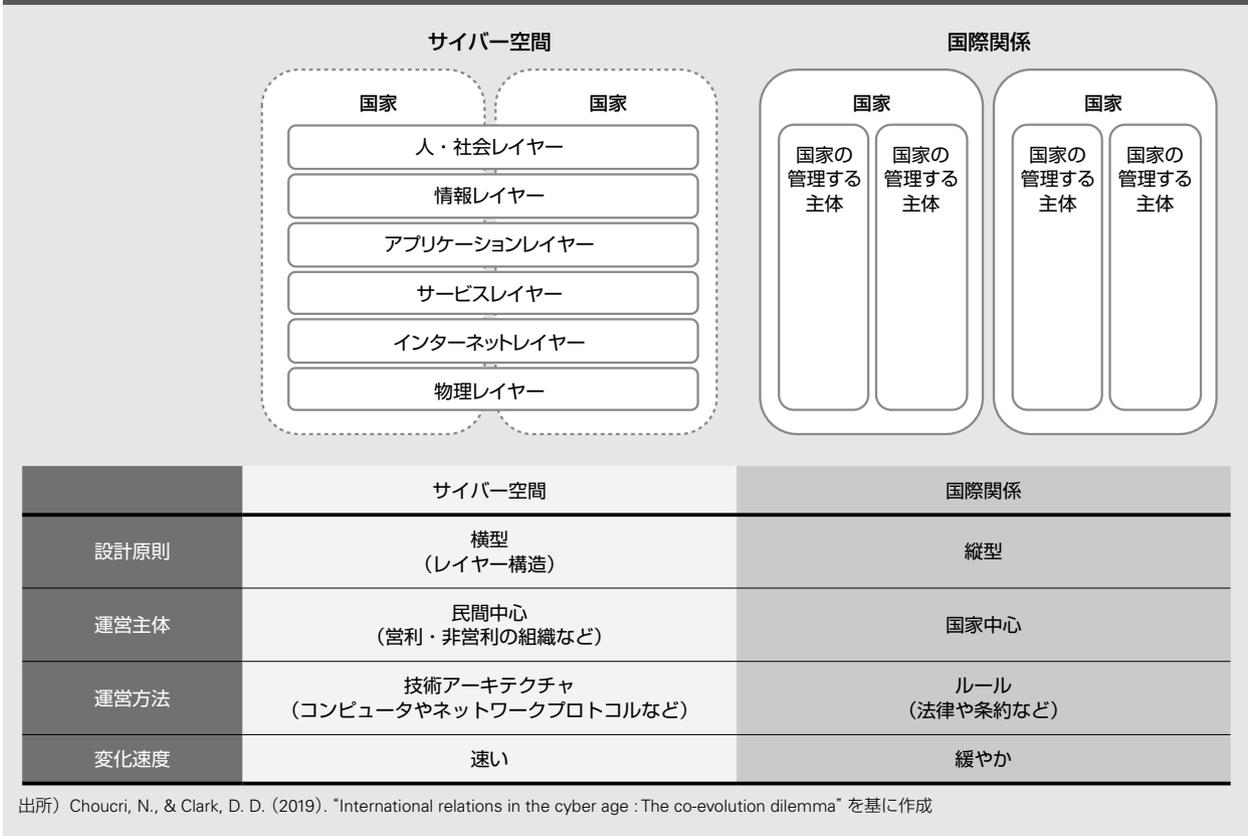
1 サイバー空間の変容

(1) サイバー空間の拡大による 社会問題の表出

国境を越えてグローバルに展開する「サイバー空間^{注3}」は、急速に強化と拡大が進んでおり、国家を基本単位とする既存の「国際関係」との不整合が課題となっている^{文献2}。両者の設計原則、運営主体、運営方法は大きく異なっており、前者は、主として民間主体がインターネットプロトコルなどの技術アーキテクチャによる運営を行う一方、後者は主として国家や国家の管理する主体が、条約や法律などのルールによる運営を行う（図1）。

全く異なる2つのシステムの不整合は摩擦を引き起こし、パワーバランスを変化させ、社会の根幹を揺るがす問題をもたらしてい

図1 サイバー空間と国際関係



る。情報漏洩やサイバー攻撃が安全保障を揺るがし、ソーシャルメディアにおける情報操作が民主主義の根幹を揺るがすなど、問題はますます拡大している。かつて、ユートピアの象徴であったサイバー空間は、今や国際政治の主要舞台の一つとなったといえるだろう。

(2) 変容するサイバー空間と多元化する社会システム

こうした事態を受けて、近年、サイバー空間に対する国家の積極的な介入が進んでいる。しかし、国家のスタンスは、多極化する世界情勢を反映して、一様ではない。冷戦後の米国一極の時代には、サイバー空間の基礎

的なルールづくりは米国が主導してきたが、多極化あるいは無極化する世界においては、各国が異なるビジョンや政策を掲げ、新たなサイバー空間と国際関係のあり方を模索している。

サイバー空間と国際関係を構成する社会システムは、こうした環境変化に伴い、異なる価値観を反映した多元的なものになっていくと考えられる。たとえば欧州はデジタル主権を掲げ、包括的なデジタル政策を展開している。半導体、データ、プラットフォーム、AI、デジタル資産など、社会システムの機能レイヤーごとに、次々と促進と規制にかかわる政策を打ち立てている。

日本も、Society5.0のビジョンの下、包括

的データ戦略に基づき、データアーキテクチャを定義したうえで、分野間データ連携基盤、ベースレジストリなどの整備が進みつつある^{文献3}。国際的にも、共通の価値観を有する国家間で、さまざまな領域における協調が進んでいる。

2 AIの飛躍的な進歩

(1) AIの飛躍的な進歩

AIの飛躍的な進歩は、このような環境下にあるサイバー空間で発生した地殻変動である。

ここ10年ほどで、データ量や計算能力の増加を背景に、機械学習^{注4}、特に深層学習^{注5}を用いて、予測や分類を行うAIが広く普及してきた。応用領域は多岐にわたり、オンラインショッピングの推薦機能、医療における病気の診断、金融における不正取引検知、自動運転など、あらゆる領域に広く普及してきた。

そうした中、文章、画像、音楽などを生成する生成AIが飛躍的な進歩を遂げ、世界に衝撃を与えた。その理由にはさまざまなものが挙げられるが、一つ選ぶとすれば、大規模言語モデルが、AI研究において象徴的な意味を持つチューリングテスト^{注6}を突破するレベルに達したことが挙げられる^{文献4}。この進展は歴史的な転換点であり、AI研究者が想定していたタイムラインを大きく塗り替えることになった。

この歴史的な進展の直接的要因は、主として、言語モデルの「べき乗則」^{注7}に基づく言語モデルの大規模化によるものであるが、その進歩の速度は驚異的である。Open AI社が2018年に発表したGPT-1のパラメータ数は1

億1700万であったが、23年に発表されたGPT-4は1兆を超えるパラメータ数を持つと考えられている。

しかし、これは通過点に過ぎない。べき乗則による成長が続くとすれば、人間の脳のシナプス数に匹敵する100兆以上のパラメータ数を持つモデルも数年以内に実現される可能性がある。また、べき乗則以外にもさまざまな方法論による基礎研究が進んでおり、AIの能力は今後もさらに進歩していこう。さらに、基礎研究を待たず、AIの社会実装は、サイバー空間のほかのさまざまな技術と融合しながら進んでいくと考えられる。既に、サイバー空間上で自律的に振舞う生成AIエージェント、生成AIを搭載したロボットなどが次々と出てきている^{文献5}。

(2) AIがもたらす社会的な影響

今後、AIが社会にもたらす影響として、生産性の飛躍的な向上、イノベーションの促進、科学的知識の発見など、さまざまな便益が期待される一方、リスクとして、雇用への影響、格差の拡大、著作権などの権利侵害、誤情報や偽情報の拡散による社会の不安定化などが議論されている。さらに、中長期的な未来のリスクとして、自律型兵器、自己複製や自己改善を繰り返すAIなどが、人類に危機的リスクをもたらすと指摘する識者もいる。サイバー空間や国際関係のパワーバランスが大きく変化していく可能性もある。

これらの想定される社会的影響を受けて、AIコミュニティの意見は、AIの研究開発を促進するべきか、抑制あるいは規制するべきかで大きく分かれている。

この対立の原因の一つは、「①AGI（汎用

人工知能)²⁸」と「②人類への危機的リスク」の実現可能性と時期に関する意見の違いである。たとえば、22人の影響力のあるAIの有識者や研究者に対するアンケート^{文獻6}では、①大規模言語モデルの成功はAGI到来のサインなのか、②もし何の対策もしなければAGIは人類に危機をもたらすのか、という2つの質問に対して、YESとNOが大きく分かれる結果となった。

もう一つの原因は、ソーシャルメディアやシリコンバレーを中心に広がりつつある、効果的加速主義と効果的利他主義と呼ばれる2つの思想の対立である。前者は、技術革新と市場の力を最大限に加速することのみ、よりよい社会変革が実現されるという考え方、後者は利他行為の効果の最大化を目指しており、そのためには技術開発の抑制や安全性の優先が正当化されるという考え方とされる。

(3) 文明の前提の再考が迫られる 社会システム

さまざまな論点や動きがあるものの、長期的な視点から見た社会システムへの最大のインパクトは、「人間はあらゆる種の中で最も優れた知能を持つ」とされてきた近代文明の前提に疑問が投げかけられ、文明の再考が迫られることにあるのではないか。「人間とAIが優れた知能を持つ」場合に、どのような社会システムをつくるべきか、それが今問われている問題だと考えられる。デカルトに始まり、啓蒙時代にその礎がつくられた社会システムが根底から再考されるときが来ているといえよう。

III 現在の「機械的な社会システム」 が抱える問題

1 経済価値と社会価値の両立が 困難であるという課題

現在の社会システムは、これらの環境変化に対応できるだろうか。現在の社会システムは、合理的に意思決定をする個人を前提としつつ、システム全体としては、与えられた目標を最大化（または最小化）する「①最適化」、法律などのルールでシステムを管理する「②他律的なガバナンス」により特徴づけられる。近代以降、人工的につくり上げられてきた社会システムという意味で、これを「機械的な社会システム（Machine-Inspired Institutions）」と呼ぶこととする。機械的な社会システムは、以下のような構造的問題があり、システム全体として、経済価値と社会価値の両立が困難であるという課題を抱えている。

(1) 最適化が生み出す外部不経済

1つ目は、最適化が外部不経済を生み出す課題である。たとえば、これまで、国はGDPの最大化、企業は株主資本の最大化、ソーシャルメディアはアテンションの最大化²⁹を目的関数として運営されてきた。その結果、われわれはより豊かになり、世界中の人とつながることができるようになったが、同時に、環境問題、格差の拡大、社会の分断などの外部不経済をつくり出してきたことも事実である。つまり、機械的な社会システムは、経済価値の創出には大きく貢献するが、一方で社会価値を棄損してしまうことが分かってきた。

(2) 他律的なガバナンスの不安定性

2つ目は、これらの外部不経済に対するガバナンスが難しいという課題である。外部不経済の発生をできるだけ抑制すべく、SDGs、ステークホルダー資本主義、テック企業への規制などの取り組みが世界中で進んでいる。しかし、環境、金融、デジタルなど、グローバルまたは国際的な問題になればなるほどステークホルダーが増え、利害関係が複雑化することで、ガバナンスはより困難になる。特にテクノロジーの場合、変化の速度が速くガバナンスが困難となる。

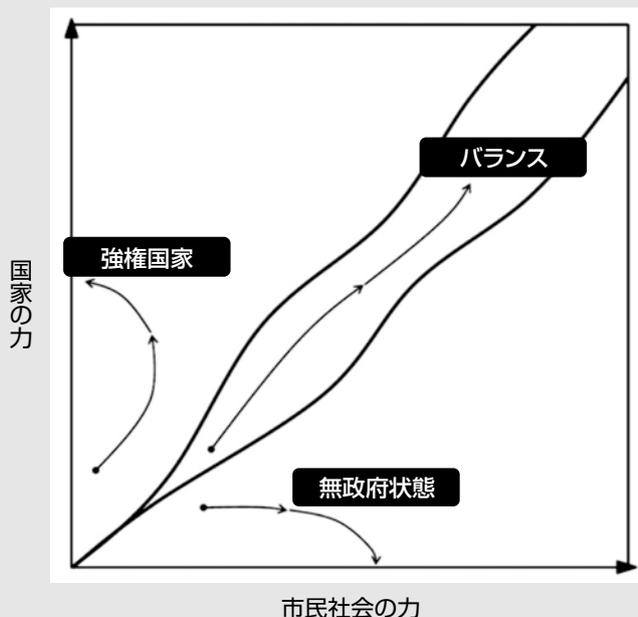
ガバナンスの困難性に関して、経済学者のAcemogluは、狭い回廊という概念を提唱している(図2)文献7。この図は、国家と市民社会のパワーのバランスを表したもので、右下に行くとも国家の力が弱くなり無政府状態となり、左上に行くとも国家の力が強くなり強権的な国家となることを表している。右下と左上への引力が強く働いており、右上のパワーバランスが取れた状態に到達することが難しいため、右上に至る狭い道を「狭い回廊」と表現している。

このことはデジタルでも同じである(図3)文献8。右下に行くとも一部の民間企業の独占となり、左上に行くとも監視国家となるため、右上の民主的なガバナンスに到達することが難しい。これは、国内のガバナンスについて述べたものであるが、国際レベル、グローバルレベルに延長して考えても、無政府状態の国家、民主国家、強権国家間の合意形成が難しいことは、容易に想像できる。

2 課題を増幅させるテクノロジー

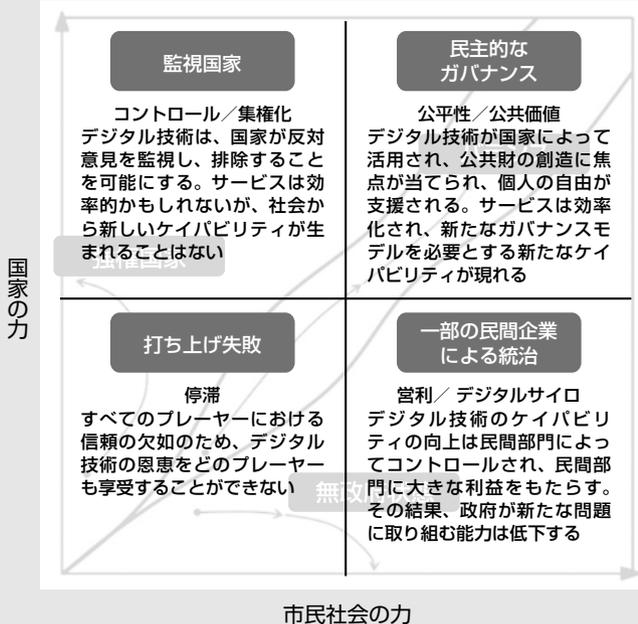
現在の社会システムを拡張するために、

図2 狭い回廊



出所) Acemoglu, D., & Robinson, J. A. (2019). "The Narrow Corridor: States, Societies, and the Fate of Liberty" を基に作成

図3 狭い回廊 (デジタル版)



出所) Eaves, D. (2022). "The Narrow Corridor and the Future of Digital Government" を基に作成

AIを中心とするテクノロジーの開発や社会実装を進めることは、経済価値と社会価値の両立が難しいという問題をさらに増幅し、危機的なリスクを現実化してしまう可能性がある。また、結果として、人間や環境中心ではなく、AI中心の世界を実現してしまうリスクがある。

(1) 最適化AIによる外部不経済発生への増幅

1つ目のリスクは、最適化を行うAIが外部不経済を引き起こすことである。例として、哲学者のBostromは、以下の思考実験を紹介している。

「紙を挟むクリップをできるだけたくさん作ることだけを目標とするAIがいたとしよう。AIは、人間がいないほうがずっといいことにすぐに気づくだろう。なぜなら、人間がスイッチを切ることを決めるかもしれないからだ。人間がそうすれば、クリップの数が減るからだ。また、人間の体にはクリップにできる原子がたくさん含まれている。AIが向かおうとしている未来は、クリップはたくさんあるが人間はいないという未来だろう」^{文献9}

クリップの製造数の最大化という最も素朴な目的関数が、人の存在を許さないとする最悪の外部不経済を引き起こし得ることを表現している点がポイントである。極端な思考実験と捉えられるかもしれないが、これまでの社会システムの歴史を振り返れば、杞憂で片づけられる問題ではないかもしれない。

計算機科学者のRusselは、現在のAI研究における「標準的なAI (Standard AI)」は、AI自身の目的達成に最適化されており、さ

まざまな問題を引き起こすリスクがあることを指摘している。また、標準的なAIは、与えられた目的が完全に正しいことが保証されている場合のみ機能するが、その仮定は現実的ではないと指摘する。解決策として「人間に有益であることが証明可能なAI (Provably Beneficial AI)」を提唱しており、そのAIは、①人間の便益の最大化という目的のみを持つ、②何が人間の便益であるかは与えられないので分からない、③人間の行動を観測することで何が人間の便益であるかを学習していく、という3つの原則に従うべきだとしている^{文献10}。

このような提案を含め、AIを人間の意図する目的や価値観と整合させていくAIアライメントという分野の研究が注目を集めている。しかし、AIの研究と社会実装のスピードに対してAIアライメントが追いつくかは定かではない。

(2) AIによるガバナンスの不安定性への増幅

2つ目は、ガバナンスが不安定になるというリスクである。DeepMind社の創業者であるSuleymanは、AIと合成生物学を中心とするテクノロジーは、前例のない規模の社会的便益をもたらす可能性がある一方で、波のように広がり、予測や制御が効かず、予期せぬ甚大な社会的悪影響をもたらすリスクがあるとしている^{注10文献11}。そして、これらのテクノロジーを安全に扱うことは、先の「狭い回廊」と同様、不安定な均衡の上に成り立っており、積極的かつ継続的な対策を要すると述べている。そのうえで、具体的対策として、技術的安全性や監査などの技術的対策、開発者や事業者や政府などの各ステークホルダー

に求められる対策、文化などの社会的対策を挙げている。

国際政治学者のBremmerはSuleymanとともに、AIの社会的影響や進歩速度の特異性を踏まえると、これまでの枠組みではガバナンスが困難であり、新たなグローバルなAIガバナンスレジーム（体制）の構築が差し迫った課題であると述べている。その対策として、テック企業を交渉のテーブルに招いた上で、気候変動、兵器拡散、金融安定の既存のグローバルなガバナンスレジームを参考にしつつ、①「AIがもたらすリスクについて事実を確定し、各国政府に助言を与えるためのレジーム」、②「全面的な軍拡競争を防ぐためのレジーム」、③「世界がかつて経験したことのないテクノロジーの破壊的な力を管理するレジーム、」という少なくとも3つのレジームを構築すべきであると提唱している^{文献12}。

実際、AIガバナンスに関する国際的な協調は、2019年のOECDのAI原則^{文献13}をはじめとしてこれまでも進んできたが、生成AIの登場以降、さらにこの動きが加速している。23年5月に開催されたG7広島サミットでは「広島AIプロセス」が創設され、10月にはAI開発組織向けの国際指針と行動規範が公表された^{文献14}。11月に英国で開催されたAI安全性サミットでは、日米中を含む28カ国の間で、AIの安全性確保に向けた国際的な取り組み強化の必要性が合意され、「ブレッチリー宣言」が公開された^{文献15}。また、欧州では6月にAI規則案が採択され、24年以降に施行される見込みであるほか、欧州評議会では、世界初のAI条約の起草の交渉が進んでいる^{文献16}。

IV 「機械と生物から学ぶ社会システム」の提案

1 経済価値と社会価値を両立させる新たな社会システム構築の必要性

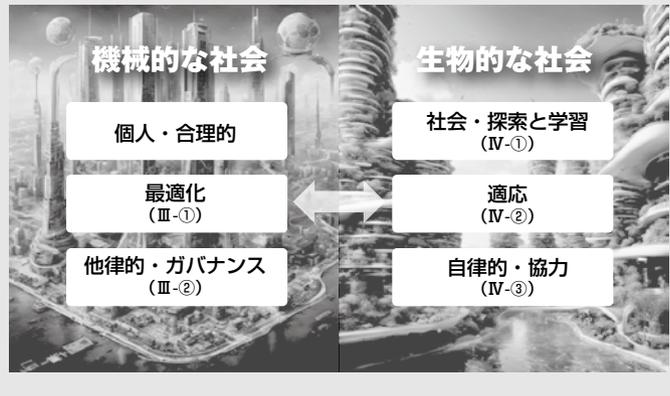
上記のように、AIが社会システムに危機的なリスクをもたらすため、それに対するアライメントやガバナンスが必要であるという議論が、現在、多くなされている。現在の社会システムが、機械的な社会システム中心である以上、このような議論になることは自然であり、安全保障や国際関係の観点からも、不可避かつ重要な議論であろう。しかし、社会システムが根底から再考を求められる中、現在の機械的な社会システムの枠組みの議論で十分だろうか。AIを中心とする強力なテクノロジーがさらに進歩する未来において、機械的な社会システムが抱える問題を克服しながら、経済価値と社会価値の両立と価値向上を可能にする社会システムを構想することこそが、今必要ではないだろうか。

本稿では、そのような社会システムとして、生物から学ぶ「生物的な社会システム (Bio-Inspired Institutions)」を取り入れた「機械と生物から学ぶ社会システム (Machine and Bio Inspired Institutions)」を提案する。

2 「生物的な社会システム」の持つ3つの特徴

生物から学ぶ生物的な社会システムは、機械的な社会システムとは対照的に、周りから学んで意思決定を行う「①探索と学習」、環境変化に柔軟に対応する「②適応」、社会課題解決に向けて社会のメンバーが利他的に協

図4 機械的な社会システムと生物的な社会システムの比較



力をする「③自律的な協力」という3つの特徴を持つ（図4）。これらの特徴を活かすことで、生物的な社会システムは、システム全体として経済価値と社会価値を両立できる。より具体的には、「①探索と学習」は経済価値を創出し、「②適応」「③自律的な協力」は集合知を高め、社会課題解決力を高めることで、社会価値の創出が期待できる。その理由について以下に示したい。

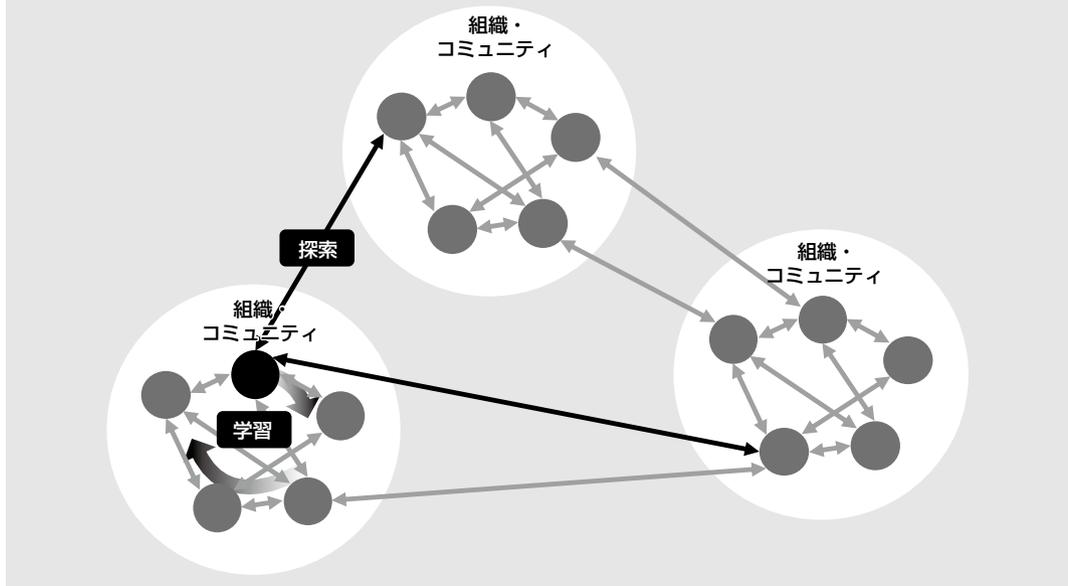
(1) 「探索と学習」による経済価値の創出

1つ目の特徴は「探索と学習」である。探索と学習は、多くの生物の採食行動において観察される。一例として、蜂の行動を紹介しよう。働き蜂は、新しい餌の場所を「探索」する。餌場を見つけた蜂は、巣に帰り、ほかの蜂に餌の場所を伝える。巣の中の蜂は、新たな餌の場所を「学習」して、少しずつ新たな餌場に向かうようになる。

生物の場合は餌であったが、人間の場合、アイデアで同じことをしていると確認されている。たとえば、企業であれば、海外や地方への出張を通じて、所属するコミュニティと異なるコミュニティに新たなアイデアを「探索」しに行く。アイデアを持ち帰り、所属コミュニティのメンバーに知らせることで、メンバーは新たなアイデアを「学習」する（図5）。

近年のビッグデータに基づく計算社会科学の研究では、より探索活動の多い人、つま

図5 探索と学習のイメージ



り、いろいろな人に会ったり、いろいろな場所に行ったりする人ほど、収入が高くなることが示されている。これは、探索活動が、より多くの機会やアイデアを提供するためだと考えられる。同様に、探索活動が多い地域や都市ほど、より経済が成長することも示されている^{文献17}。つまり、探索活動を促進するほど、経済価値が創出されることが示唆されている。

(2) 「適応」による集合知の向上

2つ目の特徴は「適応」である。たとえば、魚の群れは、外敵を察知して、自在に群れの形を変えることで生存率を向上させる。同じように、脳のニューロンは、外部の刺激に対して、柔軟にニューロンのつながりの強さを変えることで、さまざまな複雑なタスクに対応している。近年の研究では、このように集団や集合体がパフォーマンスを発揮する条件として、①フィードバック、②可塑性^{註11}が重要だと考えられている（図6）。

このことは人間社会でも同様である。シンプルなタスクを、グループのメンバーに繰り返し実施してもらう実験では、フィードバ

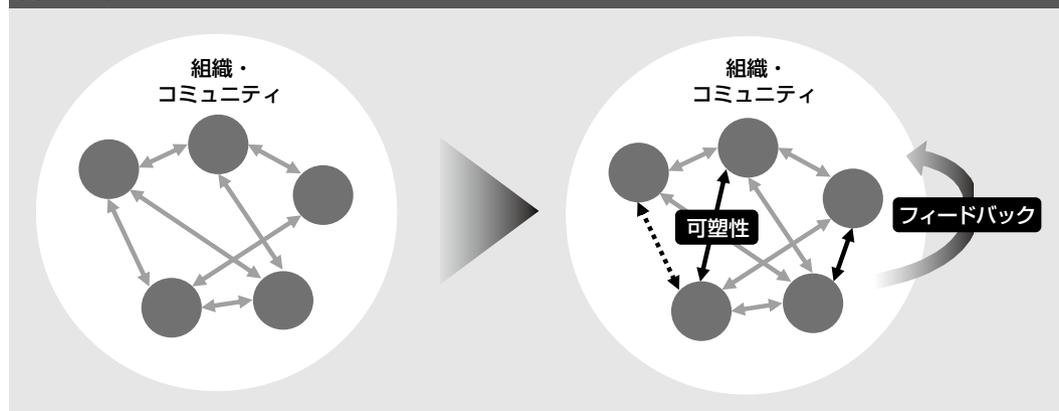
クのレベル、可塑性のレベルを上げていくと、個人とグループのパフォーマンスが向上することが示されている^{文献18}。適応を取り入れることにより、先に挙げた、環境問題、格差の拡大、社会の分断など、最適化が生み出す外部不経済を回避しつつ、社会がより賢くなる効果が期待される。

(3) 「自律的な協力」による

社会課題解決力の向上

3つ目の特徴は「自律的な協力」である。ミーアキャットや一部の鳥類などの動物は、共同で子育てを行うことが知られている^{文献19}。進化論上は、利己的に振る舞う個体の方が、より食料や資源を得ることで生存確率を高め、自然選択の過程で有利になりそうに思えるが、実際には利他的な協力的行動が多くの生物に見られる。この謎を解き明かす研究は、進化生物学などにおいて、利他性や協力の研究として知られている。ある研究では、人間はほかの種に比べて、高い社会的認知能力のあることが示されており^{文献20}、この卓越した能力による協力こそが人間の繁栄の理由との見方もある。

図6 適応のイメージ



現在われわれは、グローバルレベル、国内レベルでさまざまな社会課題に直面しており、協力の重要性は一層増している。社会課題解決のための協力の促進方法に関して、さまざまなフィールド実験をサーベイした結果では、自分や他者の貢献を見えるようにする「社会的な介入」は、金銭を渡すなどの方法で協力を求める「経済的な介入」に比べて一貫して効果が高く、社会問題の解決に役立つことが示されている^{文献21}。つまり、協力を効果的に促進することで、社会課題の解決に寄与することが期待できる。

3 問題解決を促進するテクノロジー

機械的な社会システムにおいては、社会システムの拡張機として機能するテクノロジーは、問題をさらに増幅しかねないことを説明した。一方、生物的な社会システムを取り入れた場合、社会システムの拡張機として機能するテクノロジーは、経済価値と社会価値の両立を促進する強大な力となる。

また、機械的な社会システムが、AI中心の世界をつくり出ししてしまうリスクがあるのに対して、生物的な社会システムにおいては、AIは人と人の中に入る仲介役として機能し、人間や環境に対する外部不経済を抑制

できるため、人間と環境が中心となることが期待できる。以下にその具体例を示す。

(1) 探索と学習を促進する

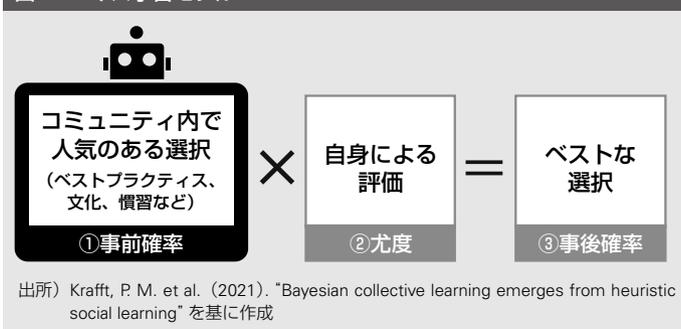
「組織・コミュニティAI」

1つ目は、探索と学習を促進する「組織・コミュニティAI」である。組織・コミュニティは、国、地方自治体、企業、市民団体など、十分な学習データが確保できる規模であれば、どのような単位でもかまわない。組織・コミュニティ単位に、生成AIや大規模言語モデルをつくることで、探索と学習の促進が期待される。

その理由は、生物的な社会システムでは、周りから学習して意思決定をしているためである。最近の研究では、組織・コミュニティにおける個人の意思決定は、ベイズ学習モデルにより説明できることが示されている^{文献22}。このモデルでは、①まず「組織・コミュニティ内で人気のある選択」をピックアップする（事前確率）、②その選択に対して自身の評価を掛け合わせる（尤度）、③その結果、ベストな選択を得る（事後確率）、という3段階で意思決定が進む（図7）。「組織・コミュニティ内で人気のある選択」とは、企業であればベストプラクティス、国や地域であれば文化や慣習などを指す。

生成AIが、与えられた学習データの範囲から統計的な常識を生成するものだとすれば、組織・コミュニティのデータを学習させることで、「組織・コミュニティ内で人気のある選択」に相当する情報を生成できると考えられる。その結果、この情報にアクセスするコストを下げるのと同時に精度を上げることで、探索と学習が促進され、経済価値の創出

図7 ベイズ学習モデル



が期待できる。

加えて、既に「組織・コミュニティ内で人気のある選択」を、経験を通じて獲得している人よりも、いまだ獲得していない、コミュニティの新規参加者ほど恩恵が大きいと考えられる。たとえば、企業であればベテラン社員よりも新入社員や中途採用社員、国や地域であれば長く居住している人よりも新規移住者に恩恵があると考えられる。よって、組織・コミュニティ内における、スキルや能力の格差を縮小する方向に働くことが期待される。

(2) 適応を促進する「ファシリテーションAI」

2つ目は、適応を促進するファシリテーションAIである。事例としてCrowdSmart AI^{文献23}を紹介する。たとえば、ある地域で空き地ができた場合に、公園をつくるか、市民会館をつくるかを市民で決めたいとする。CrowdSmart AIは、そんなときに、関係する市民に公平にヒアリングしているいろいろな意見を整理・集約したうえで議論の方向性を提案するなど、ファシリテーターとしての機能を提供してくれる。

AIファシリテーションの利点は、人間と異なり、公平に意見をヒアリングできること、ファシリテーションに偏りがでないこと、プライバシーを守れることなどにある。社会の適応力をAIが高めることにより、組織やコミュニティは、より賢い意思決定ができるようになる。

(3) 協力を促進する「コーポラティブAI」

3つ目のアイデアは自律的な協力を促進する「コーポラティブAI」である。Cooperative AI Foundation^{文献24}は、今後の社会にお

いて、人とAI、AIとAI、AIと組織などの新たな関係性が生まれることを想定したうえで、社会厚生を高めることを目的に、協力する能力を持つAIの研究を進めている。協力を理解するためには、さまざまな分野の知見を統合する必要がある。たとえば、協力を構成する要素として、他者の意図を汲み取る能力や共感力などの能力、喜びや怒りなどの感情が挙げられる。社会的なメカニズムで協力を促すものとしては、互惠性、評判、社会規範、制度などの要素がある。

直近のNeurIPS2023では、Melting Potと呼ばれるシミュレーション環境において、効果的に協力を促進するためのアイデアを競うコンテストが開催されている^{文献25}。シミュレーション環境では、放っておくとエージェントが利己的に振る舞い、食料や資源を枯渇させる「共有地の悲劇」と呼ばれる問題が発生する。この問題に対処するために、協力的なエージェントを導入して、どのような能力や条件を持たせることが効果的かを探ることがコンテストの目的である。

(4) 適応と協力を促進する

「フィードバックAI」

適応と協力はフィードバックや可視化によっても促進される。これを可能にするのが「フィードバックAI」である。組織・コミュニティ内の統計情報や活動レベルを可視化してフィードバックするシステムは、既に多くのアイデアが考案または社会実装されているが、AIを取り入れることで、より双方向的な対話や高度なデータ分析が可能になり、フィードバックや可視化の効果をより高めることが期待できる。

たとえば、カーボン排出量をフィードバックして、脱炭素に向けた協力を促進することができる^{文献26}。自身や組織・コミュニティ内の人と人のつながり（社会関係資本）を、プライバシーに配慮しながら可視化してフィードバックすることで、ウェルビーイングの向上に役立てることもできる^{文献27}。地域において、いろいろな人に会ったり、いろいろな場所に行ったりするなど、探索活動のレベルを可視化してフィードバックすることで、よりよい都市計画や意思決定に活用すること

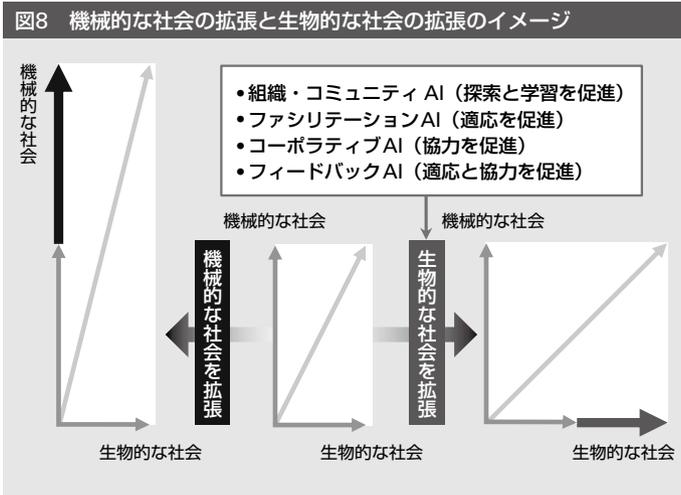
もできるだろう^{文献28}。

4 経済価値と社会価値の両立と価値向上を可能にする

「機械と生物から学ぶ社会システム」

機械的な社会システムは、経済価値を創出するが社会価値を棄損してしまう。一方で、生物的な社会システムは、経済価値と社会価値の両方を創出できる。したがって、現在の機械的な社会システムに、生物的な社会システムを取り入れた「機械と生物から学ぶ社会システム」は、経済価値と社会価値の両立と価値向上が期待できる。変化の速いテクノロジーが引き起こす問題やリスクに対しても迅速かつ柔軟な対処が可能となる。

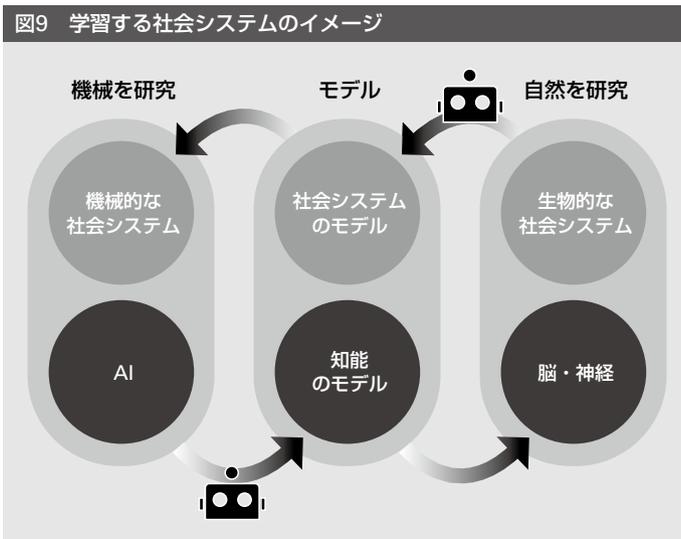
また、社会システムの拡張機であるテクノロジーを、機械的な社会システムではなく、生物的な社会システムの拡張に利用することで、機械と生物から学ぶ社会システムに、より効率的に近づくことができるだろう（図8）。



5 持続的な価値向上を可能にする

「機械と生物から学習し続ける社会システム」

機械と生物から学ぶ社会システムのさらに未来の姿として、経済価値と社会価値の持続的な価値向上を可能にする「機械と生物から学習し続ける社会システム」が構想できる。AI、ロボット、人工生命などの歴史を振り返ると、自然や生物から学びモデルを構築し、それを基に人工物や機械をつくり、その挙動を学びモデルを更新する、という学習サイクルを繰り返してきた。同様に社会システムも、生物と機械の両方から学習するサイク



ルをつくれれば、継続的によりよい社会システムのモデルを更新していくことができるはずである。つまり、「機械的な社会」と「生物的社会」に関するわれわれの理解が継続的に更新されていくということである。

さらに、この学習サイクルにおいて、AIが知識の発見を支援することで、サイクルの加速が期待できる。AIが、よりよい資本主義や民主主義のあり方を見つけて、社会科学を進展させる未来は近いかもしれない（図9）。

V 日本への示唆とステークホルダーに求められる視点

機械的な社会システムに進みつつある世界が、機械と生物から学ぶ社会システムへ転換していくうえで、日本の文化が果たす役割は大きい。「三方良し」「空気を読む」などで表される日本の文化は、生物的社会システムとの親和性が高いからである。これらの文化は、われわれが歴史を通じて獲得してきた知恵である。多元的な世界においては、情報に還元できない複雑な文化や価値観こそが最大の強みとなる。これまで、日本は生物的社会システムから機械的な社会システムへ転換することを進めてきたが、あらためて自分たちの持つ文化や価値観を振り返ることが求められているのかもしれない。

新たな社会システムを構築するうえでは、市民が新たな社会システムを実験的に創っていくことが鍵となる。既存の社会システムを構成する国や地方自治体や企業などには、制度整備、支援、協働などを通じて、市民の新たな社会システム創りを促し、そのケイパビ

リティを取り込み、自らを変革する役割が期待される。

萌芽事例は世界中で既に多く出てきている。RadicalxChangeは、多元的で民主的な未来社会に資する技術を開発して、世界各国で社会実装を進めている^{文献29}。欧州では、資金提供プログラムやハッカソンなどを通じて、戦略的にCivic Techを育成する動きが見られる。日本では、市民主体のCode for Japanが東京都とのコロナのダッシュボードの開発を始め、国や地方自治体や企業と協力して、数々の成果を上げている。

このような市民中心の取り組みが今後さらに広がることを期待したい。本稿が次の社会システム創りにかかわる方々にとっての考える材料を提供できたとすれば幸いである。

注

- 1 ここていう「社会システム」とは、国、企業、ソーシャルメディアなどの、人と人が協調するための仕組みを指す。英語の「Institutions」を指すが、一般的な訳語である「制度」は、税制など個別の制度の意味合いが強いため、「社会システム」と呼ぶこととした。なお学術的には「Institutions」にはさまざまな定義がある。たとえば、経済学者のNorthは「政治的、経済的、社会的相互作用を構造化する、人間が考案した制約である」としている
- 2 AIなどの研究領域では、テクノロジーを「人間を置き換える（replace）道具」としてではなく「人間の能力を拡張（Extension）する道具」と捉える考え方がある。この考え方を応用して、本稿ではAIを中心とするテクノロジーを「社会システムの能力を拡張する道具」と位置づけることとする
- 3 国際政治学者のChoucriとインターネットプロトコルのアーキテクトであるClarkは「サイバー空

間は、人間が相互作用するグローバルな領域であり、①インターネット上の多数のコンピュータの相互接続によりつくり、②（コンピュータなどの物理的構成要素で、論理的な相互接続の枠組みを可能にする）レイヤー構造で構築され、③情報の処理・操作・利用・増強、人と情報の相互作用を可能にし、④制度的な仲介と組織によって利用できるようになり、⑤分権性と、参加者や関連するグループや利害関係者間の相互作用によって特徴づけられる」としている

- 4 データから学習して予測などを行う手法のこと
- 5 機械学習の一つで、ニューラルネットワークを多層化してデータの特徴量を自動的に学習する手法のこと
- 6 コンピュータが人間相当の知能を持つかを評価するテストであり、評価者が壁越しにいるコンピュータと会話をし、それがコンピュータだと見分けられなければ、人間相当の知能があると見なされる。実際のAI研究において必ずしも広く使われるテストではないが、象徴的な出来事として、ここでは取り上げた
- 7 ①訓練データの量、②訓練に投入する計算量、③利用するモデルのパラメータ数、という3つの変数と予測誤差の間にべき乗則が成り立つという法則。2020年にOpen AIが論文で発表した
- 8 人間と同様の汎用的な知能を持つ人工知能という意味で使われることが多いが、国際的に合意された定義は本稿執筆時点では存在しない
- 9 広告のクリック率などを最大化すること
- 10 Suleymanはこの問題を「Containment Problem」と呼んでいる
- 11 柔軟に個人間のつながりを変えられるとともにつながりの形状を記憶する性質のこと

参考文献

- 1 North, D. (1991) "Institutions", *Journal of Economic Perspectives*
- 2 Choucri, N., & Clark, D. D. (2019) "International relations in the cyber age: The co-evolution dilemma", MIT Press
- 3 https://www.digital.go.jp/policies/data_strategy/
- 4 Biever, C. (2023) "ChatGPT broke the Turing test—the race is on for new ways to assess AI" <https://www.nature.com/articles/d41586-023-02361-7>
- 5 Wang, L, et al. (2023) "A Survey on Large Language Model based Autonomous Agents"
- 6 Strickland, E., & Zorpette, G. (2023) "The AI Apocalypse: A Scorecard", *IEEE Spectrum* <https://spectrum.ieee.org/artificial-general-intelligence>
- 7 Acemoglu, D., & Robinson, J. A. (2019) "The Narrow Corridor: States, Societies, and the Fate of Liberty", Penguin Press
- 8 Eaves, D. (2022) "The Narrow Corridor and the Future of Digital Government", Ash Center for Democratic Governance and Innovation, Harvard Kennedy School <https://ash.harvard.edu/narrow-corridor-and-future-digital-government>
- 9 Miles, K. (2014) "Artificial Intelligence May Doom The Human Race Within A Century, Oxford Professor Says", *Huffington Post* https://www.huffpost.com/entry/artificial-intelligence-oxford_n_5689858
- 10 Russel, S. (2019) "Human Compatible: Artificial Intelligence and the Problem of Control", Penguin Books
- 11 Suleyman, M. (2023) "The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma", Crown
- 12 Bremmer, I., & Suleyman, M. (2023) "The AI Power Paradox : Can States Learn to Govern Artificial Intelligence—Before It's Too Late?", *Foreign Affairs*
- 13 <https://oecd.ai/en/ai-principles>
- 14 https://www.mofa.go.jp/mofaj/ecm/ec/page5_000483.html
- 15 <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/>

- the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023
- 16 <https://ifi.u-tokyo.ac.jp/project-news/16287/>
- 17 Chong, S. et al. (2020) “Economic outcomes predicted by diversity in cities”, EPJ Data Science
- 18 Almaatouq, A. et al. (2020) “Adaptive social networks promote the wisdom of crowds”, Proceedings of the National Academy of Sciences
- 19 Balter, M. (2014) “Human altruism traces back to the origins of humanity”, Science
<https://www.science.org/content/article/human-altruism-traces-back-origins-humanity>
- 20 Esther Herrmann et al. (2007) “Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis”, Science 317, 1360–1366
- 21 Kraft-Todd et al. (2015) “Promoting Cooperation in the Field”, Current Opinion in Behavioral Sciences
- 22 Krafft, P. M. et al. (2021) “Bayesian collective learning emerges from heuristic social learning”
- 23 <https://www.crowdsmart.ai/>
- 24 <https://www.cooperativeai.com/>
- 25 <https://assets.aicrowd.com/challenges/meltin-gpot-challenge-2023>
- 26 <https://www.code4japan.org/activity/jibun-goto-planet>
- 27 Nishikata, T. et al. (2019) “Social Capital Accounting, Trusted Data: A New Framework for Identity and Data Sharing”, MIT Press
- 28 <https://inequality.media.mit.edu/>
- 29 <https://www.radicalxchange.org/media/blog/plurality-technology-for-collaborative-diversity-and-democracy/>

著者

西片健郎（にしかたたけお）

野村総合研究所（NRI）未来創発センターデジタル社会研究室エキスパートリサーチャー

専門はデジタル社会インフラの技術アーキテクチャの研究開発、国際標準化