ワークショップ報告書
CRDS-FY2022-WR-05

科学技術未来戦略ワークショップ報告書

トラスト研究戦略

～デジタル社会における新たなトラスト形成～

CRDS 研究開発戦略センター
Center for Research and Development Strategy

国立研究開発法人科学技術振興機構

エグゼクティブサマリー

本報告書は、2022年6月11日に開催した科学技術未来戦略ワークショップ「トラスト研究戦略～デジタル社会における新たなトラスト形成～」の内容をまとめたものである。

トラスト（信頼）は、相手が期待を裏切らないと思える状態だと考えられる。リスクがあるとしても、相手をトラストできると、安心して迅速に行動・意思決定ができる。顔が見える身近な人たちの間で育まれた「旧来のトラスト」は、デジタル化の進展に伴い、うまく機能しなくなってきた。バーチャルな空間にも人間関係が広がり、複雑な技術を用いたシステムに依存するようになり、情報の非対称性が拡大している。さらに、だます技術も高度化している。このような状況に対して、デジタル社会においてもうまく機能する新しいトラストの仕組み作りが必要になってきた。

このような問題意識のもと、JST CRDSでは2021年6月から「デジタル社会における新たなトラスト形成」をテーマとして、人文・社会科学から人工知能、医療まで、幅広い分野の動向を俯瞰するセミナーシリーズやワークショップ、さまざまな分野・立場の有識者へのインタビューなどを通して調査・検討を進めてきた。本ワークショップでは、その検討結果に基づき、JST CRDSから、本テーマの目指す姿と意義、研究開発の方向性に関する素案を示すとともに、さまざまな立場の有識者14名に参加をお願いし、それぞれの切り口から話題提供、コメント、議論をいただいた。

トラストには、「対象真正性」（本人・本物であるか?）、「内容真実性」（内容が事実・真実であるか?）、「振る舞い予想・対応可能性」（対象の振る舞いに対して想定・対応できるか?）という3側面があり、現在の取り組みの多くは、このいずれかの側面に重点を置いている。今後の方向性として、これら3側面を併せてトラストを多面的・複合的に扱っていくことの必要性が高まると考えている。

本ワークショップでの話題提供も、このトラストの3側面を踏まえて、対象真正性に重点を置いたトラストサービス、内容真実性に重点を置いたフェイク問題、振る舞い予想・対応可能性に重点を置いたTrustworthy AI（信頼される人工知能）のそれぞれの研究開発状況・課題・方向性を紹介いただいた。さらに、これらに横串を通す法制度・ガバナンスの考え方・方向性についても話題提供をいただいた。

それを受けて、産業界の視点、および、人文・社会科学の視点から、取り組みの方向性や意義、学際的研究の推進方法などを含めて、コメンテーターからの意見を皮切りにして総合討議を行った。デジタル社会における新たなトラストの仕組みは、技術開発・制度設計による対策だけでなく、産業的なインパクトや社会受容性といった面からも検討していくことが必要なことを再確認した。

科学技術振興機構（JST）研究開発戦略センター（CRDS）は、科学技術に求められる社会的・経済的なニーズを踏まえて、国として重点的に推進すべき研究領域や課題、その推進方策に関する提言を行っている。本ワークショップもその一環として開催したものであり、これまでの調査・検討結果や本ワークショップでの議論に基づき、「デジタル社会における新たなトラスト形成」を推進するための戦略プロポーザルの発行[1]を予定している。

1　2022年9月末頃に公表予定「戦略プロポーザル：デジタル社会における新たなトラスト形成」（CRDS-FY2022-SP-03）

目次

1 ｜ 問題意識および提言の方向性 …… 1

1.1 問題意識 …… 1

1.2 戦略提言の方向性 …… 5

2 ｜ 技術開発・制度設計の状況・課題・方向性 …… 9

2.1 対象真正性の視点から
手塚 悟 …… 9

2.2 内容真実性の視点から
藤代 裕之 …… 16

2.3 振る舞い予想・対応可能性の視点から
杉村 領一 …… 21

2.4 法制度・ガバナンスの視点から
稲谷 龍彦 …… 28

3 ｜ 社会・産業および人文・社会科学の視点から …… 33

3.1 産業界の視点から
浦川 伸一 …… 33

3.2 人文・社会科学の視点から (1)
川名 晋史 …… 36

3.3 人文・社会科学の視点から (2)
内田 由紀子 …… 39

4 ｜ 総合討議 …… 42

付録　ワークショップ開催概要 …… 51

第２章・第３章の発表者の所属等は以下の通り。
なお、第４章の議論も含めた参加者リストは付録に記載。

発表者名	所属等
手塚 悟	慶應義塾大学環境情報学部 教授
藤代 裕之	法政大学社会学部 教授
杉村 領一	産業技術総合研究所 情報・人間工学領域 上席イノベーションコーディネータ
稲谷 龍彦	京都大学大学院法学研究科 教授
浦川 伸一	損害保険ジャパン株式会社 取締役専務執行役員
川名 晋史	東京工業大学リベラルアーツ研究教育院 准教授
内田 由紀子	京都大学 人と社会の未来研究院 教授

1 ｜ 問題意識および提言の方向性

1.1 問題意識

トラスト（信頼）は、相手が期待を裏切らないと思える状態だと考えられる。リスクがあるとしても、相手をトラストできると、安心して迅速に行動・意思決定ができる。顔が見える身近な人たちの間で育まれた「旧来のトラスト」は、デジタル化の進展に伴い、うまく機能しなくなってきた。バーチャルな空間にも人間関係が広がり、複雑な技術を用いたシステムに依存するようになり、情報の非対称性が拡大している。さらに、だます技術も高度化している。このような現状認識から、デジタル社会においてもうまく機能する新しいトラストの仕組みが必要だというのが、本日の議論の出発点になる問題意識である（図1–1）。

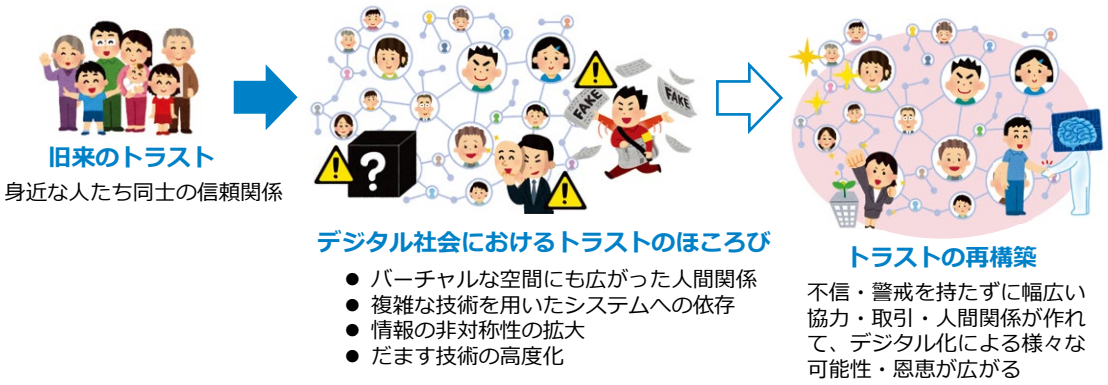


図1–1　　デジタル社会におけるトラストの問題

デジタル化の進展が生んだトラスト問題の具体的なシーンを図1–2に挙げた。ネットを介して人間関係が広がるとともに、デジタル技術を用いただます技術が高度化し、「仮想世界のトラストに基づく取引」における偽装・なりすましや「メディアにおけるフェイク拡散」が社会問題化した。人工知能（AI）などの複雑な技術を用いたシステムへの依存が進んでいるが、そのようなブラックボックス的な技術を用いたシステム「自動運転車」「パーソナルAIエージェント」「人を評価するAIシステム」をトラストできるかという問題もある。さらに、トラスト関係は、ネットを介した人間関係に加えて、人間とAI・ロボットとの関係にも広がり、「医療意思決定におけるAIセカンドオピニオン」「コミュニケーションロボット」「メタバース内活動におけるトラスト」など、新たな様相を呈しつつある。

トラストの役割・効果として、ニクラス・ルーマンは「社会的な複雑性の縮減メカニズム」[1]と捉えた。工藤郁子は「取引や協力のコストを減らしてくれる社会関係資本」[2]、レイチェル・ボッツマンは「確実なものと不確

1　ニクラス・ルーマン著、大庭健・正村俊之訳、『信頼：社会的な複雑性の縮減メカニズム』、勁草書房、1990年。

2　工藤郁子、「人々の「眠り」と「目覚め」、社会の信頼 再構築を」、朝日新聞デジタル「にじいろの議」2020年8月12日。https://www.asahi.com/articles/DA3S14584996.html

実なものの隙間を埋める力であり、未知のものに自信を持つこと」[3]と書いている。また、トラストはビジネスを発展させ、かつ、ビジネス上の競争要因である。取引・協力できる相手を拡大し、先端技術を用いたビジネスを加速し、ビジネス上の意思決定の迅速化・不安低減を可能にし、消費者・取引参加者に安心をもたらす。

仮想世界のトラストに基づく取引

ネットの世界で、リアルには面識のない人たちとの取引や、仮想通貨・デジタル資産を用いた取引が拡大。さまざまな分野で、シェアリングビジネス、不特定多数の中からのマッチングビジネスも立ち上がっている。

仮想世界・デジタルデータの性質を悪用した偽装やなりすまし等の犯罪も起きている。対策も取られているが、常に新しい仕組みが生まれ、新しいリスクが発生し、対策が追いつかない面もある。

メディアにおけるフェイク拡散

最新のAI技術によって人間の認識能力では見破れないフェイクの作成が容易になった。ソーシャルメディアの普及によって、フェイクやデマの拡散が大規模化。

フェイク動画による政治干渉や個人攻撃・棄損等が社会問題化。簡単に人を騙すことができてしまい、裁判等での証拠の信頼性も揺らぐ。フェイクの法的規制が強くなると、表現の自由が妨げられる恐れも生じる。

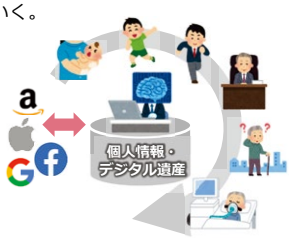
自動運転車

AI技術を活用した状況認識や運転制御によって、車が運転手なしで走行する地区・状況が広がる。

AI技術はブラックボックスで動作保証や精度保証ができない。
安心して乗車できるのか？
事故が発生したときに、原因解明や責任の所在はどうなるのか？

パーソナルAIエージェント

個人情報の管理代行をパーソナルAIエージェントに任せるサービスの利用が増えていく。



ブラックボックスで、個人の意図・期待の通りに振る舞うという100%保証はできないのに、個人情報を委ねることができるか？
期待に反する事態が起きた場合の責任の所在はどうなるのか？

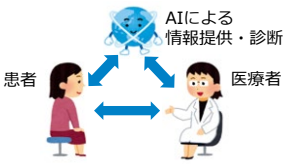
人を評価するAIシステム

採用試験の一次フィルタリング、人事評価や配属最適化といった人事業務へもAIシステムが応用されつつある。
中国では個人の様々な行動履歴を追跡し、信用スコアを算出して、優遇や制限を与えるシステムが稼働している。

学習データに偏りや差別的要因が含まれていると、不公平で差別的な評価を助長するという問題がある。また、評価アルゴリズムに過剰適合して行動する人々を生み出す。

医療意思決定における AIセカンドオピニオン

従来は患者と医療者の二者関係による医療意思決定だった。そこに、AIによる情報提供や診断が加わり、三者関係による医療意思決定へと変化しつつある。



患者が異なる多様な情報を参照できるようになり、医療者からの説明と異なる情報が得られることもある。三者関係における新たな役割関係とそこでのトラストのあり方が必要になっている。

コミュニケーションロボット

親近感を持てる外観や、会話を含むコミュニケーション能力を備えたコンピューターエージェントあるいはロボットが、さまざまなサービスや日常的なタスクにおいて、対人インターフェースとして広がる。

人間らしい外観を持っていると、人間並みの能力を持っていると期待・過信してしまい、そうでないと失望するといったことが起きやすい。その一方、身近なロボットに、過度な親近感・依存感を持ってしまうタイプの人もいる。

メタバース内活動に おけるトラスト

仮想世界の中で自分のアバターを介した新たな人間関係や経済活動が生まれる。

生身の人間や物理的な実体が必ずしも確認できない世界において、リアル世界と同様のトラストが成り立ち得るか？

図1–2　デジタル化の進展が生んだトラスト問題

3　レイチェル・ボッツマン著、関美和訳、『TRUST：世界最先端の企業はいかに〈信頼〉を攻略したか』、日経BP社、2018年。

既に昨年（2021年）の俯瞰セミナーシリーズと俯瞰ワークショップ[4]で幅広く紹介・議論されたように、トラストにはさまざまな定義やモデルがあるが、ここでは議論をしやすくするために、おおむね共通の理解に近いと思われる整理を図1–3に示した。特にリスクと主観という2点はポイントであろう。「Trusted Webホワイトペーパー」を参考にして二重円で表現したが、裏付けのある部分と裏付けがない部分があって、それはリスクがある状況でトラストするということになる。また、どの程度リスクがある状況でトラストするかは、Trustorの主観に依存する。それから、詳しい説明は後で行うが、トラストについて、対象真正性、内容真実性、振る舞い予想・対応可能性という3側面に分けて考える。

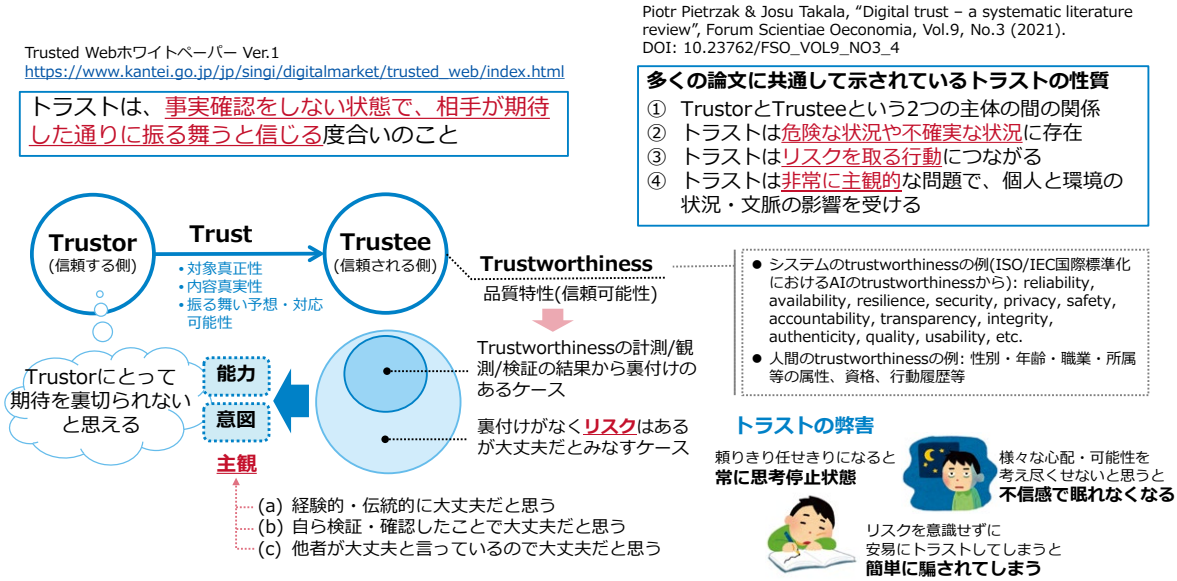


図1–3　　トラストに関する整理

図1–4　　トラストに関する研究開発の状況

4　科学技術振興機構研究開発戦略センター、「俯瞰セミナー＆ワークショップ報告書：トラスト研究の潮流　～人文・社会科学から人工知能、医療まで～」、CRDS-FY2021-WR-05、2022年2月。
https://www.jst.go.jp/crds/report/CRDS-FY2021-WR-05.html

昨年（2021年）の俯瞰セミナーシリーズと俯瞰ワークショップでさまざまな話題が論じられたように、トラスト研究は、人文・社会科学分野の基礎研究から、ビジネスや社会実装と連動した情報科学分野の技術開発、医療やSNS（Social Networking Service）などの応用シーンでの対策検討まで、幅広く取り組まれている（図1–4）。特にセキュリティー・データ戦略、AI戦略・ガバナンスで戦略的な取り組みが進められている。

1.2 戦略提言の方向性

ここから、トラスト研究戦略に関して、JST CRDSで検討してきた方向性や提言骨子について話していく。

さまざまな分野でトラストに関わる研究が実施されていることは先ほど話した通りだが、それらの間で知見共有・連携はほとんどない状況にある。その結果、それぞれはトラスト問題に対して個別的な対処や断片的な状況改善にとどまっている。

そこで、図1–5のBeforeからAfterのような形に変えたい。つまり、デジタル社会における新しいトラストの仕組みとそれによるトラスト問題対策の全体ビジョンを描いて共有し、具体的トラスト問題と共通基礎の両面から連携して、社会に貢献する研究を目指すのが望ましい。

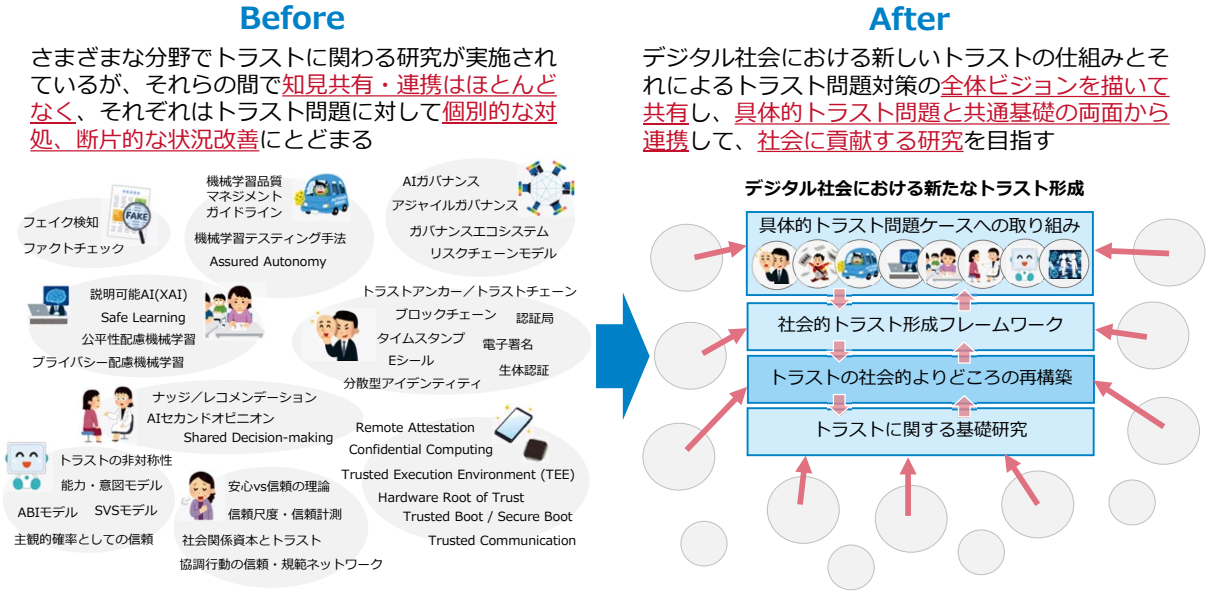


図1–5　　トラスト研究の目指す姿

このような変容に際して中核となるのが「トラストの社会的よりどころの再構築」だと考える。

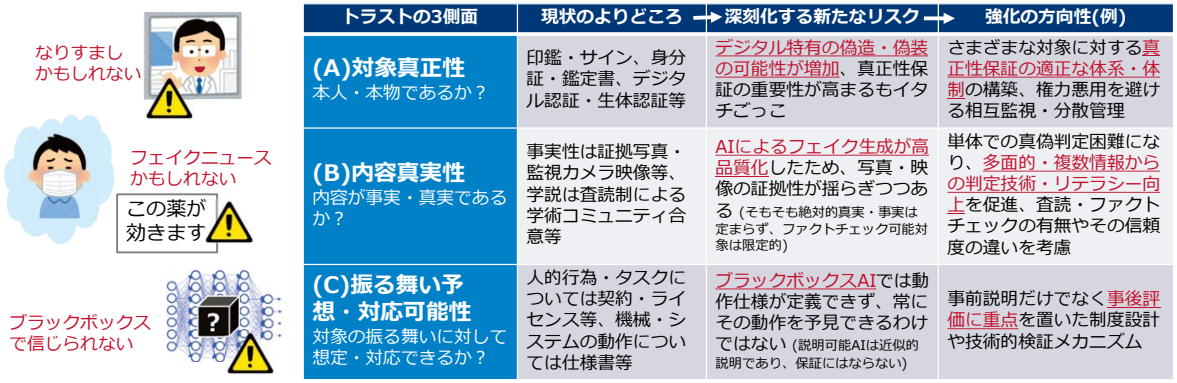
これに関連して、先ほども少し触れた「トラストの3側面」、すなわち、(A)「本人・本物であるか?」という対象真正性、(B)「内容が事実・真実であるか?」という内容真実性、(C)「対象の振る舞いに対して想定・対応できるか?」という振る舞い予想・対応可能性がある。例えば、相手がなりすましかもしれないというのは対象真正性に関する疑念、フェイクニュースかもしれないというのは内容真実性に関する疑念、システムがブラックボックスで信じられないというのは振る舞い予想・対応可能性に関する疑念である。

現在のトラスト研究がばらばらに見えるのは、トラストの3側面の異なる側面を扱っていることが、一つの原因と思われる。つまり、デジタルトラストの技術開発は対象真正性、フェイク対策の技術開発は内容真実性、AIのトラストに関する技術開発は振る舞い予想・対応可能性にフォーカスしている。これからの方向性として、それら3側面を統合した多面的・複合的な枠組みへの再構築が望まれる。

また、既に述べたように、トラストするか否かは最終的にTrustorの主観に依存する。しかし、詐欺のような犯罪を防ぎ、社会秩序を維持するため、社会的よりどころになるものが用意される。現状でも、身分証、鑑定書、デジタル認証、証拠写真、監視カメラ映像、契約書、仕様書などが、各側面の社会的よりどころと

して使われている。

ところが、デジタル化の進展に伴い、偽装・偽造の可能性が増え、AIによるフェイク生成も容易になり、ブラックボックスAIの応用が広がったことで、従来の社会的よりどころだけでは不十分になってしまった（図1–6）。それゆえ「トラストの社会的よりどころの再構築」が必要なのである。



トラストの3側面	現状のよりどころ	→ 深刻化する新たなリスク	→ 強化の方向性(例)
(A)対象真正性 本人・本物であるか？	印鑑・サイン、身分証・鑑定書、デジタル認証・生体認証等	デジタル特有の偽造・偽装の可能性が増加、真正性保証の重要性が高まるもイタチごっこ	さまざまな対象に対する真正性保証の適正な体系・体制の構築、権力悪用を避ける相互監視・分散管理
(B)内容真実性 内容が事実・真実であるか？	事実性は証拠写真・監視カメラ映像等、学説は査読制による学術コミュニティ合意等	AIによるフェイク生成が高品質化したため、写真・映像の証拠性が揺らぎつつある（そもそも絶対的真実・事実は定まらず、ファクトチェック可能対象は限定的）	単体での真偽判定困難になり、多面的・複数情報からの判定技術・リテラシー向上を促進、査読・ファクトチェックの有無やその信頼度の違いを考慮
(C)振る舞い予想・対応可能性 対象の振る舞いに対して想定・対応できるか？	人的行為・タスクについては契約・ライセンス等、機械・システムの動作については仕様書等	ブラックボックスAIでは動作仕様が定義できず、常にその動作を予見できるわけではない（説明可能AIは近似的説明であり、保証にはならない）	事前説明だけでなく事後評価に重点を置いた制度設計や技術的検証メカニズム

図1–6　　トラストの3側面と社会的よりどころの課題

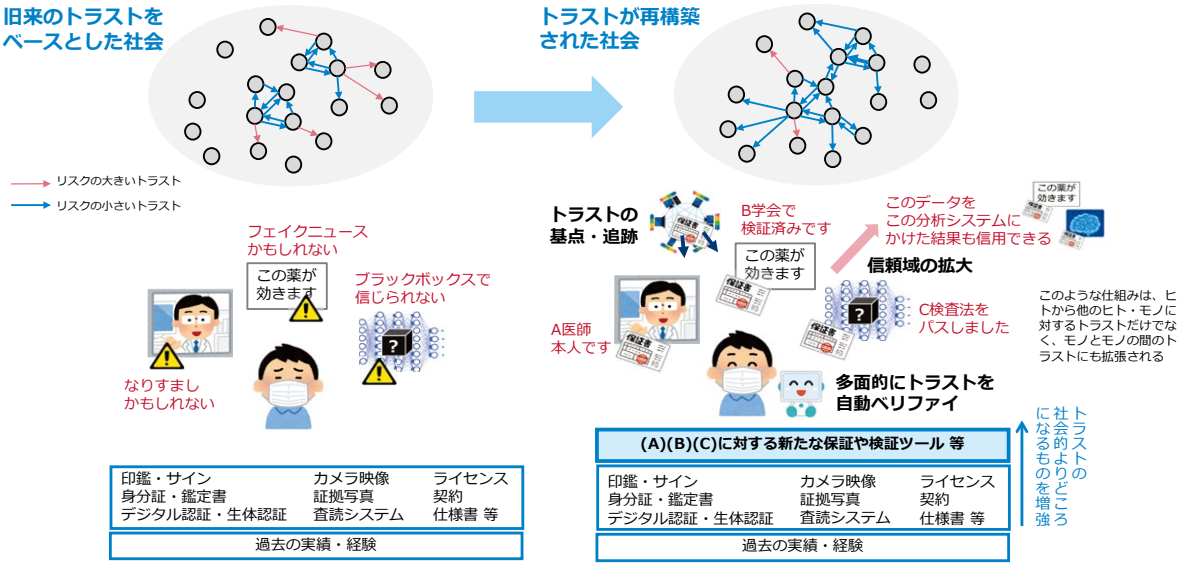


図1–7　　トラストの社会的よりどころの再構築（Before/Afterのイメージ図）

図1–7は、トラストの3側面に対する社会的よりどころのBefore/Afterのイメージを描いたものである。Afterでは、トラストの3側面に対する新たな保証や検証ツールなどを用意することで、トラストの社会的よりどころになるものを増強する。それによって、Beforeでは「なりすましかもしれない」「フェイクニュースかもしれない」「ブラックボックスで信じられない」といった、トラストの各側面での疑念が生じていたのに対して、Afterでは新たな判断材料が与えられるとともに、そのような判断材料を複数集めて、多面的・複合的にベリファイ（Verify）することを容易にする手段も提供される。それを用いて、トラストできる対象範囲を広げて

いくことも可能になる。また、トラストの基点になるものが、権力者や特定勢力に支配されないように、分散監視などの機能も設けていく。

このようなトラストの社会的よりどころの再構築に基づいた新たなトラストの仕組みが、社会に根付き、さまざまなトラスト問題への対策に結びつくようにするための研究開発課題を、図1-8に示す4層で構成する。すなわち、第1層「トラストの社会的よりどころの再構築」を中核として、それを社会に広げるための第2層「社会的トラスト形成フレームワーク」、さらに、それらを適用した第3層「具体的トラスト問題ケースへの取り組み」を推進する。また、第1層・第2層・第3層が社会に受容され、人々に浸透していくために、人文・社会科学研究を中心とした第0層「トラストに関する基礎研究」も重要である。このような全体ビジョンを共有し、文理分野横断で社会実装と共通基礎を連携させて推進していくことが必要と考える。

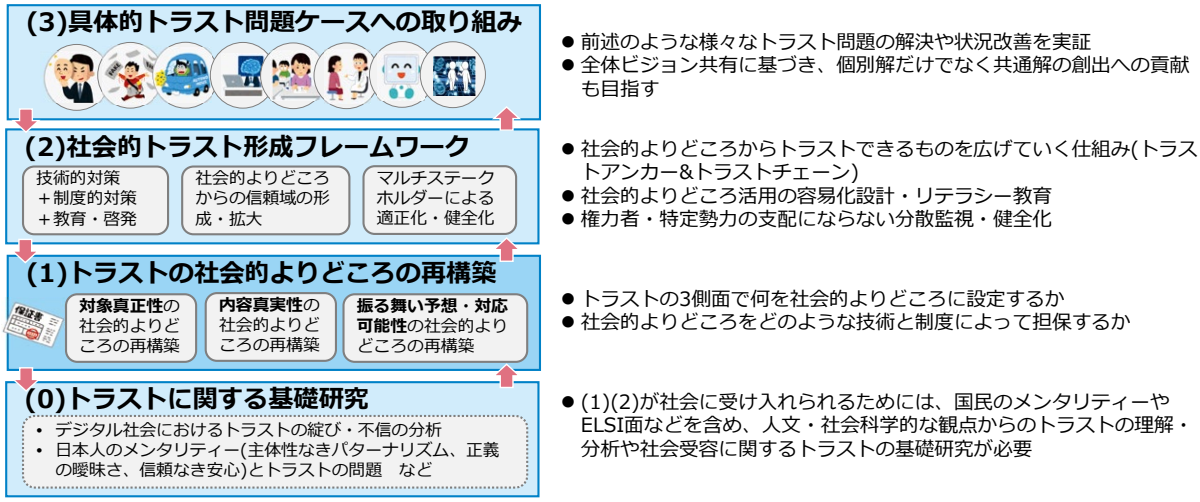


図1-8　　4層から成る研究開発課題（案）

最後に、このような研究開発を推進・加速するための施策についても考えており、意見・示唆をいただきたい。議論のための素案として図1-9を示す。この図の中央の青色の部分は、研究コミュニティーや研究体制がこのような方向に発展できるとよいというイメージを示した。これに対して、外からの緑色の矢印は政策・施策として考えたものの一案である。

全体ビジョンを共有した分野横断・総合的なトラスト研究を推進する体制への変容のため、分野間の知見共有・連携の場作りが有効と考える。昨年（2021年）に実施したセミナーシリーズやワークショップを、クローズドな形態ではなくオープンな形態で実施することを考えている。そこからさらに、さまざまな研究分野に波及するトラスト共通基礎研究の育成のため、トラスト共通基礎研究の拠点立ち上げが考えられるかもしれない。

また、そのような分野横断の議論が進んで、全体ビジョンを踏まえた具体的な目標を設定できるようになったら、それを推進するためにファンディングプログラムを活用することが考えられる。図にも示したように、現在、トラストの3側面やガバナンス面で、既に戦略的な活動が動いており、そこからさらに発展する形で全体ビジョンにつながるシナリオも考えられる。そのような取り組みが、日本がトラストを軸とした技術戦略で国際的にリードするポジション強化にもつながり得ると期待する。

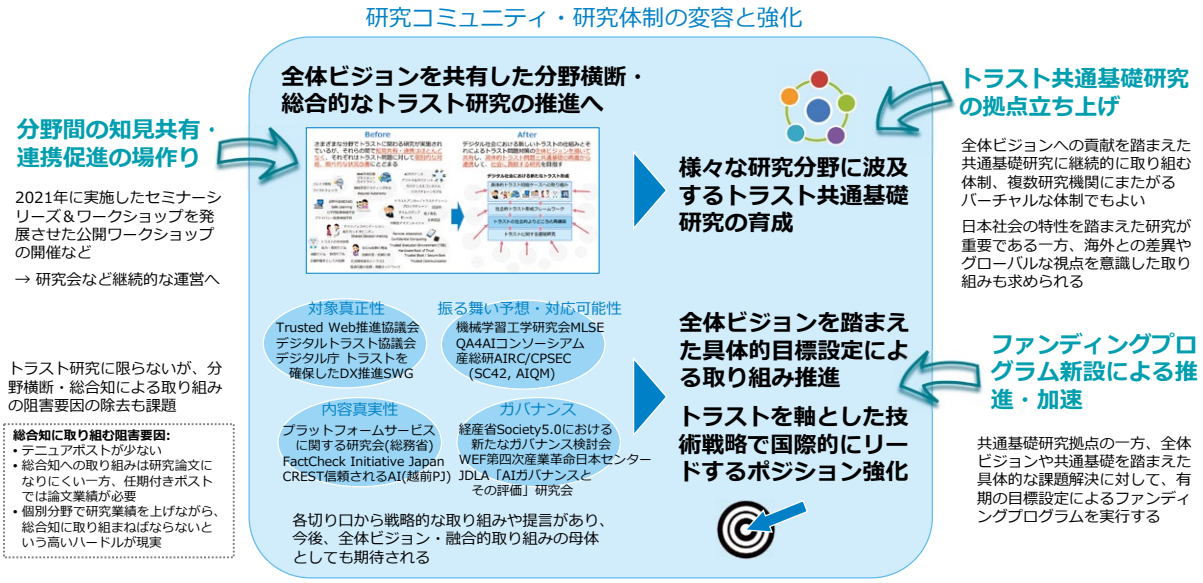


図1–9　　推進方策（議論のための素案）

以降、トラストの3側面と法制度・ガバナンスの視点から、取り組みの状況・課題・方向性について、4名から発表いただき（第2章）、それを受けて、産業界と人文・社会科学の視点から3名のコメンテーターにコメントをいただき（第3章）、最後に総合討議を行う（第4章）。

なお、第2章以降、発表資料から引用した図の権利は発表者に属し、了解のもと引用させていただいた。文中の人名は敬称を省略させていただいたほか、文責はJST CRDSにある。また、本報告書中に記載したURLは2022年7月末時点のものである。

2 ｜ 技術開発・制度設計の状況・課題・方向性

2.1 対象真正性の視点から

手塚 悟

デジタルトラストという観点から、対象真正性について説明する。デジタルトラストは、Data Free Flow with Trust（DFFT）でいうと“T”の部分にあたり、デジタル庁では、「データ戦略推進ワーキンググループ」の配下にデジタルトラストを専門的に検討する「トラストを確保したDX推進サブワーキンググループ」を設置して議論している。今回は、そこでの議論の内容と今後の課題を中心に説明する。

2.1.1 トラストサービスの意義

Society 5.0の中核となるデータ駆動型社会（Data-driven Society）やデジタルトランスフォーメーション（DX）では、良質、最新、正確かつ豊富なリアルデータが価値の源泉となり、経済社会活動を支える最も重要な糧となることが見込まれる。これは、とりもなおさず、経済社会を支える中核的な要素としてのデータの重要性が飛躍的に増大することを意味する。このようなさまざまな可能性を秘めるデータ駆動型社会においては、そのバックボーンとなるデータの真正性やデータ流通基盤の真正性を保証することが極めて大切となる。そのためには、インターネット上におけるヒト・組織・モノ・データ等の真正性・真実性・正当性を確認し、データの改ざんや送信元のなりすまし等を防止する仕組み（トラストサービス）の実現が必要となってくる。

ワールドワイドでみると、EUはデジタルシングルマーケットを創設するために、その基盤を支える包括的なトラストサービスの法制化を進めている。また、国連のUNCITRA（United Nations Commission on International Trade Law）という商法系の機関から「Identity Management and Trust Services」というモデル法が発表される予定と聞いている。

2.1.2 トラストサービス

トラストサービスの機能は、電子署名やタイムスタンプ、eシール、ウェブサイト認証、モノの正当性の認証、eデリバリーがある（図2-1-1）。

また、トラストサービスに関係する用語として、情報セキュリティーの7要素があり、基本3要素の「機密性」「完全性」「可用性」に加えて、プラス4要素として「真正性」「信頼性」「責任追及性」「否認防止」がある。

トラストサービスに関連する用語として、AuthenticationとCertificationがあるが、どちらも日本語では「認証」と訳すため、混乱することがある。AとBが、その2人の間で認証することをAuthenticationと呼び、その時のよりどころになるのが「What you know（知識認証）」、「What you have（所有認証）」、「What you are（生体認証）」で、これらで相手を特定する。一方、Certificationについては、例えば、公開鍵方式で、秘密鍵と公開鍵のペアキーのうち、公開鍵にお墨付きを与えることをCertificationと呼ぶ。「公的個人認証」の「認証」は、もともとはCertificationの意味だが、公的個人認証の中には認証行為もあり、それはAuthenticationである。

次に、サイバー空間におけるエンティティーのIDについて説明する。エンティティーのIDはヒト・組織・モノ・データのようにエンティティーとなるもの全てについて必要である。例えば、ヒトにはマイナンバー等、組織には法人番号等、モノには個体番号等、データにはハッシュ値等がIDとなる。また、IDは、サイバー

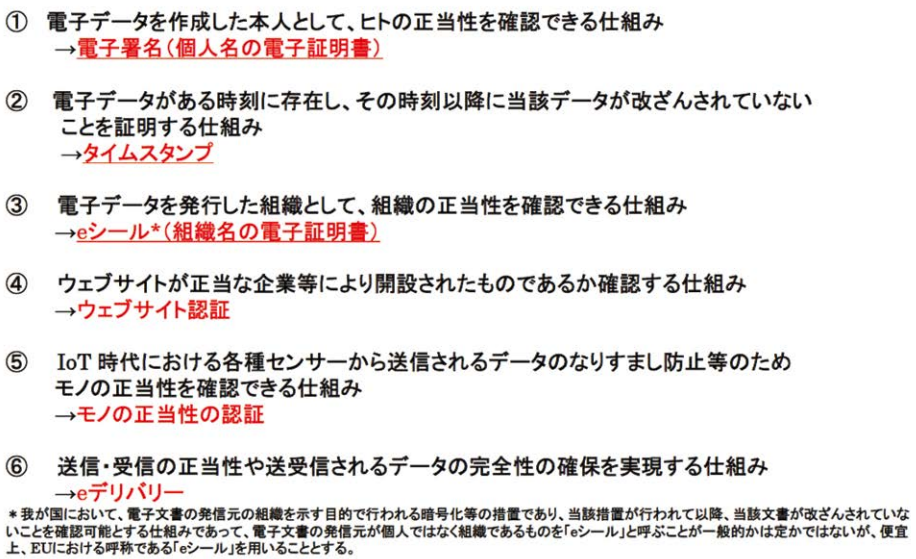


図2–1–1　　トラストサービス

空間でのエンティティーの識別と、さらに、サイバーフィジカルシステム（CPS）における、フィジカル空間で誰かを特定するためにも必要であり、この識別と特定のために、きっちりと社会インフラの中に組み込む必要がある。

あわせて、トラストアンカーの概念は、非常に重要であり、まさにトラストアンカーをどこに置くかで全くそのシステムが変わってくる。トラストアンカーは、トラストサービスの中のトラストの源である。事前にA、Bを審査・登録してIDを発行し、トラストアンカーがA、Bの存在を確認することで、データの改ざんやなりすましを防いでいる（図2–1–2）。トラストを確保したDX推進サブワーキンググループでは、トラストアンカーを政府に置くのか、民間に置くのか、個人に置くのかという、大きくは3種類について検討している。法制度と紐付けると、トラストアンカーは政府に置くであろうし、民間協会などのグループ内でトラストサービスを提供する場合は、民間の管理するトラストアンカーは民間のグループの中に置き、民間が管理するであろう。トラストサービスでは、こういった概念をきっちりして設計しないといけない。

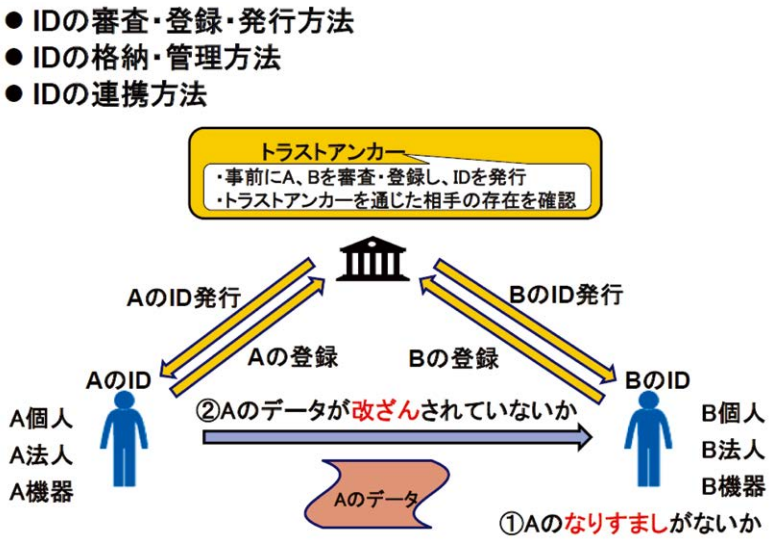


図2–1–2　　トラストアンカー

その中では、ValidationとVerificationが必要となる。A（Subscriber）からB（Relying Party）にデータを渡す場合、Bは、Aのなりすましがないかと、Aのデータが改ざんされていないか、この2点を検証できないといけない。そのためには、Aは、ローデータではなく、Validated Dataを生成し、Bは、AのValidated DataをVerification（検証）できるようにしておく必要がある。こういう社会メカニズム、インフラを作らないと、トラストがサイバー空間上で構築できないし、さらには、自動化に行き着かない。

2.1.3 トラストチェーン

トラストチェーンについては、国際的なトラストに関する相互承認の枠組みを検討しており、法制度、運用、テクノロジー、全ての点で、国際的に構築しないといけない（図2–1–3）。日本とEUでは、2022年5月12日に日EUデジタルパートナーシップが立ち上げられている。テクノロジーサイドでも、今後、この枠組みを検討していかないといけない。

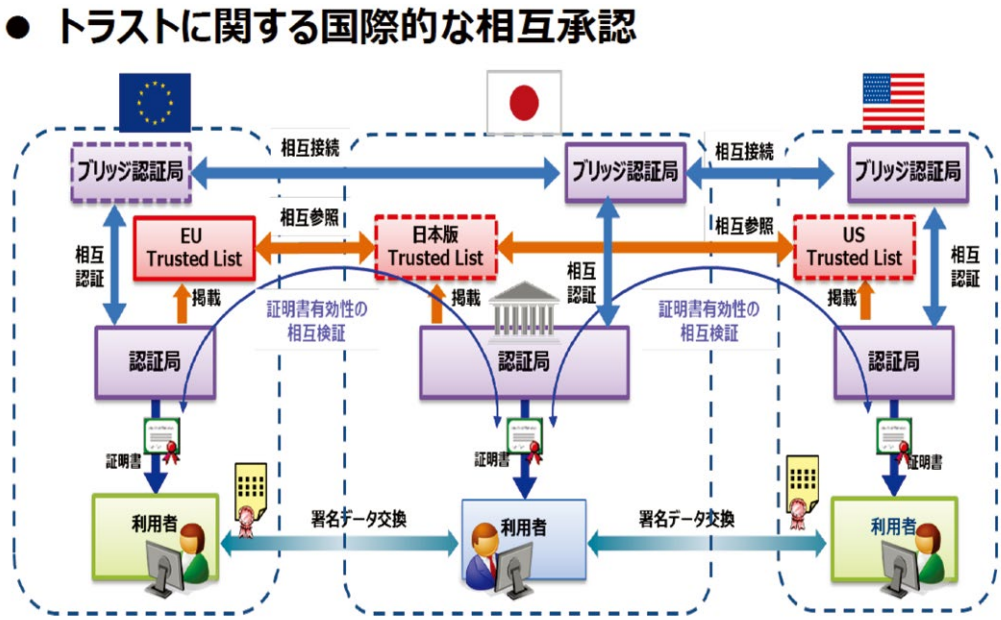


図2–1–3　　トラストチェーン

2.1.4 トラストサービスの保証レベル

トラストサービスの保証レベルは、身元確認（IAL）、クレデンシャルの強度（AAL）、トラストサービス事業者の信頼度（TAL）で決定される。また、ベースレジストリと紐付けたデジタルIDをトラストサービス事業者から発行するスキームの創設が重要となる（図2–1–4）。

トラストサービス事業者はIDプロバイダーやクラウド署名、認証局、タイムスタンプ局を含んだトラストアンカーであり、このトラストサービス事業者の信頼度（TAL：トラストサービスのアシュアランスレベル）を定義しないといけない。この辺りには、まだ研究していかないといけない内容がある。図2–1–5は、IALとAAL、TALの関係の例（電子署名法の場合）である。

また、IDプロバイダーについては、NISTのSP800–63–3と比較して、対象とするIDをどうしていくかの検討も必要である。

もう一つ重要な点は、パブリックトラストとプライベートトラストという概念である。プライベートトラストは、いろいろな自主基準に従う、例えば、民間企業グループの中でのトラストの提供である。一方、パブリックトラストは、本当にパブリックな世界でのトラストで、法制度や国際標準などの基準に従い、非常にスケーラブルな仕組みになる。国際的な発展も含めて考えると、しっかりとした基準作りをして、スケーラブルな仕組みを構築することが重要である。

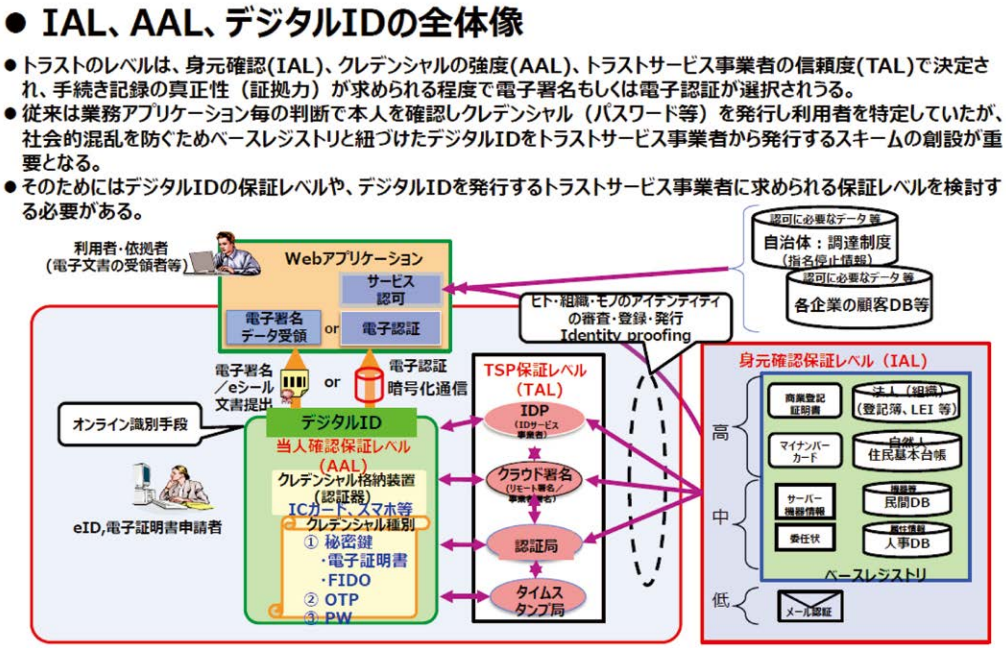


図 2-1-4　　トラストサービスプロバイダーと保証レベル

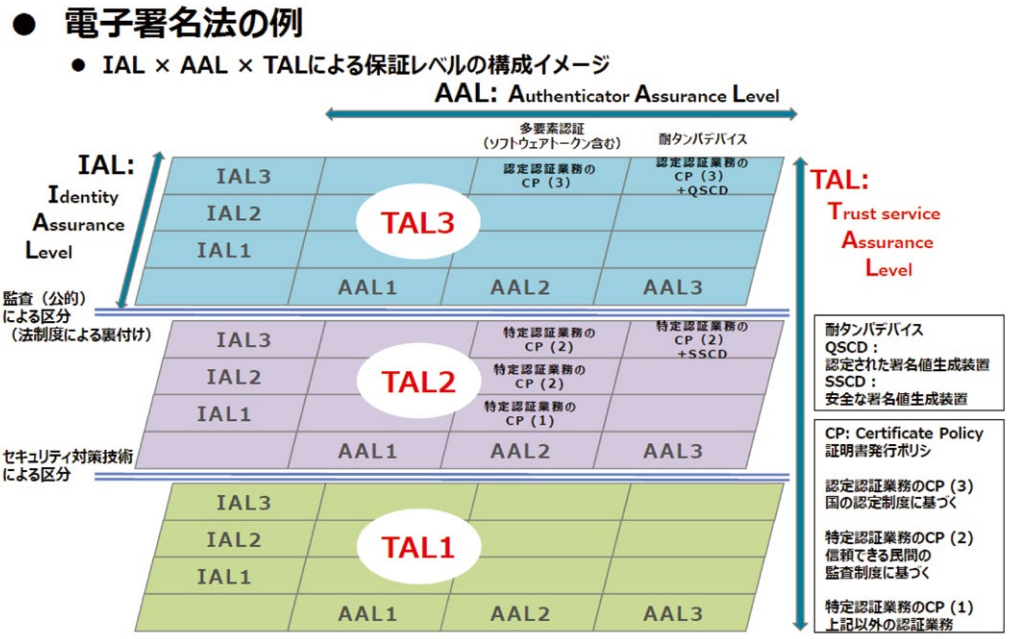


図 2-1-5　　IAL、AAL、TALの例

2.1.5 アプリケーションサービスとトラストサービス

アプリケーションサービスとトラストサービスの関係（図2-1-6）を見ると、トラストサービス層には、トラストの概念を含めて、認証局、認可局、ID管理がある。これまでは、それぞれのルールがばらばらに決められていたため、パブリックトラスト化してスケーラブルにしようとすると大きな壁があった。このため、パブリックトラストという概念で、トラストサービス層をパブリック化できるように構築していくことが非常に重要となる。パブリックトラストでは、透明性が重要になるため、そのフレームワークとして認定の枠組みが必要になる。認定の枠組みでは、トラストサービスのアシュアランスレベルを考えていくことが必要で、今はまだ概念設計の段階であり、今後検討しないといけない。

それ以外には、トラストサービスの実現のために、ヒトとデータの区分（ヒトのセキュリティークリアランスやデータのクラシフィケーション）をした上でアクセス制御を実施することや、経済安全保障と絡む国家間や重要インフラシステム間等で情報共有をする際、秘匿情報分類のデータの区分が必要となる。また、正当なヒトが正当なレベルのデータへアクセスするためには、認証（Authentication）と認可（Authorization）というアクセス制御の概念が必要であり、IDのレベルに応じて、そのレベルに応じたデータにアクセスできるようにする必要がある。Society 5.0でも、同様に、アクセス制御が必要である。

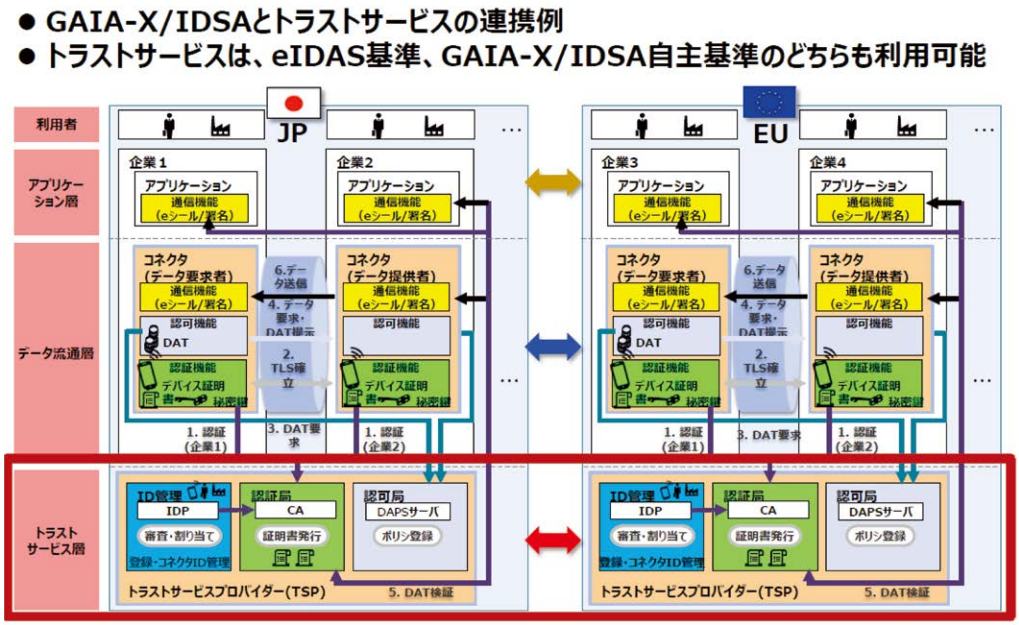


図2-1-6　　トラストサービスに関する保証レベル

2.1.6 トラストサービスの国際相互連携

最後に、国際相互連携については、「自由と信頼」のルールに基づくデータ流通圏と国際相互連携が必要である。例えば、コンテンツ、契約情報の真正情報などをやり取りする時に、なりすましと改ざんのない状態にする。そのためには、トラストサービス基盤とトラストデータ流通、トラストアプリケーションサービスの3階層の中のトラストサービス基盤が、国際間で相互連携できないといけない（図2-1-7）。例えば、自動車を動くIoT機器と考えると、その動くIoT機器を欧州でも日本でも特定できるように、トラストサービスの国際相互連携を実現していかないといけない。そのためには、法制度、監督・適合性評価、技術基準、トラスト

アンカー間の接続の仕組みという４項目の同等性などを検討し、相違点を補完することが必要で、今後研究しながら整備していく必要がある。

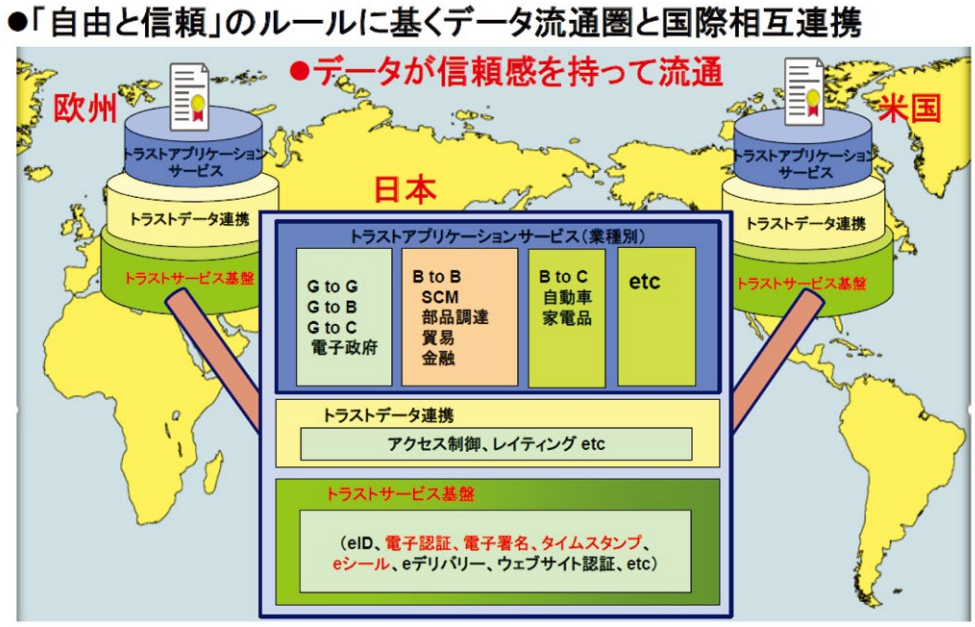


図２−１−７　　トラストサービスの国際相互連携

【質疑・討議】

中川：ヒトとヒトとの間のトラストは重要だが、それ以外にも、例えば、法人のようなもののトラストも対象になってくる。また、ヒトではなくヒトの代理のようなもの、メタバースにおけるアバターやＡＩのエージェントなどのケースもある。また、１人の人が複数のアバター使う、アバターが自律的に動くケースもあるかもしれない。自律的に動いているアバターとヒトとの間のトラストをどう考えるかなども生じる。こういった場合、ヒトとして捉えるのか、それとも、モノに対するトラストで考えていくのか、同じ仕組みを適用してしまっていいのかなど、どうこの枠組みの中で捉えていくのか、もう少し知りたい。

手塚：まず、テクノロジーサイドから言うと、対象を切り刻んでいったときに、どれだけのエンティティーで表現できるかという点がある。次に、それらのエンティティーを組み合わせて、その集合体が社会で動く一つのアプリケーションやミドルウェアになるという点がある。また、エンティティーと、複数のエンティティーの組み合わせた集合体は、分けて議論しないといけない。複数エンティティーの集合である「セットオブエンティティーズ」といった概念をもう一つ導入しないといけないと思うが、まだそういう議論がされていない。

例えば、ヒトの場合、ヒトを切り刻むというのはあり得ないが組織の場合はある。組織の場合、当然合併など、いろいろ変化があるので、そこは分かりやすい。モノは、サプライチェーンの概念による合体型だと思う。これらの議論は、今後やっていかないといけない。

人間の作業スピードが分単位なのに対して、サイバー空間はナノ単位で作業できる点がすごい。それにより、ヒトの代わりにＡＩの活用が進み、インとアウトだけは人間が関わるが、それ以外は基本的に全部サイバー空間でできる。さらに、データドリブン型の考え方で、データとデータを複数インテグレートして新たなサービスを生むようなことがサイバー空間で起こると思う。そのとき、一番プリミティブなエンティティーが何で、それにＩＤが振られていて、それらが組み合わさってまた新しいサービスが生ま

れると思う。そういう概念がまだまだ固まっていないので、そういうところが今後は重要になってくる。

福島：今回、対象真正性の側面からお話しいただいたが、他の側面との関係も今後重要になってくると思う。特に内容真実性についてはどうお考えか？　内容真実性には確からしさの度合いが入ってきて、それを複合的に見ていく必要があるので、対象真正性の枠組みに内容真実性の情報が追加されるようになるのかなど、お聞きしたい。

手塚：これについては、トラストを確保したDX推進サブワーキンググループでもかなり議論があり、内容真実性と対象真正性は定義を分けている。

内容真実性は、例えば、体温計などの温度センサーから出てきた値を信じるか信じないか、信じるとするならばなぜ信じられるのかの問題である。36.5℃と出た値を信じるかは、ちゃんとしたメーカーのブランドだから、では、そのブランドは何によって信頼ができているかというと、製造工程からすべてを含めたサプライチェーン、その中の一つ一つのプロセスの確からしさに基づいて作られていて、最終的に温度センサーの温度が誤差範囲に入っているから、我々はその値を信じていると思う。つまり、プロダクトのバックグラウンドでどういうことがされたかという属性情報、いろいろな属性の真実、その真正、確からしさが合わさって、そのプロダクトは出来上がっているから信頼できると言えるのだと思う。

今度は、内容真実性のある情報を、今のトラストサービスのように、サイバー空間でほかの人に渡す。このAからBに渡すというのは対象真正性の概念になり、36. 5℃という値が改ざんされていないことと、なりすましがないことを担保する必要がある。

内容真実性と対象真正性の2つのフェーズは、分けて議論しないといけない。まずは対象真正性の方から検討している。内容真実性は、サプライチェーンでのバックドアなどさまざまな課題を克服して初めて信頼できることになるだろうと議論している。

2.2 内容真実性の視点から

藤代 裕之

フェイクニュースを含めてニュースを中心に、インターネットメディアのトラストについて考えてみる。

2.2.1 メディアリテラシーのパラドックス

フェイクニュースに対抗するためには、メディアリテラシーを向上させようという話になりがちだが、それではトラストを担保できないという、メディアリテラシーのパラドックスについて話をしたい。

総務省や新聞通信調査会などのアンケート調査によると、既存メディア、すなわちテレビや新聞の信頼度は高く、インターネットの信頼度は低いという結果が出ている。既存メディアの信頼は高いが利用者は減少し、スマホ、タブレットなどから接触するネットやソーシャルメディアの信頼性は低いが利用者は増加している、というのが現状である。おおむね確かなメディアがあるという共通認識がトラストのよりどころであるとすれば、ネットはトラストな状態とは言えないだろう。

なお、おおむね確かなメディアの参考になるのが『信頼を考える』の第３章エスノメソドロジーにおける信頼概念にある、「同一の世界は、事実的世界によって賦課された偶発的事象のもとで、立場の交換可能性という仮定を維持する人々の能力によって保証される」（Garfinkel 1963、秋谷訳（2018））という間主観的同一性である。異なる人の間で共通の認識が成り立ち得るということがメディアのトラストの根源であると言える。しかしながら、ネットではさまざまな情報やニュースが断片化するため信頼は分裂、断片化する（成田 2015）。

内閣官房デジタル市場競争本部のTrusted Web推進協議会のホワイトペーパーの中に、信頼に関連してこういう記述がある。

「SNSで流れる情報について、元々の情報がどの程度信頼できるソースから発信されたものか、それがどのような経路をたどって、どの程度信頼できる介在者がどのように加工したのかなどが不透明であり、現状では一部のプラットフォーム事業者が付すラベリングに頼らざるを得ない点がある。」[1]

ネットの信頼は断片化しているため、ポータルサイトに掲載されているから、検索結果に表示されているから、というプラットフォーム事業者が示すラベリングに依存せざるを得ないが、ポータルサイトや検索サイトは、表示されているニュースや情報に対する責任を担保しているわけではなく、あくまで掲載しているという立場に過ぎない。つまり、ネットのトラストは根拠なきラベリングに依存しているという非常に脆弱な状況にある。

ホワイトペーパーの記述には続いて、「本来は情報の信頼性について多角的に検証できることが必要である」とある。現状ではホワイトペーパーが指摘するように、ソースや経路を多角的に検証しようとしても、ポータルサイトのニュースの編集方針や検索サイトの表示順を決めるアルゴリズムなどは提示されておらず、検証することは不可能である。その状態を改善し、多角的に検証できるようにする必要がある。だが、一部の人々は、問題がある現状を改善せずに、メディアリテラシーを推進しようとしている。

ここでトラストの３原則とされている内容真実性が大きな問題となる。フェイクニュースや陰謀論に関して、内容真実性を確認することはできない。理由はそれが嘘だからである。嘘をいくら確認しようとしても嘘であり、これがフェイクニュースや陰謀論の決定的な強さである。この決定的な強さに対して、既存メディアを批判的に読み解けば、間違いを見つけ出すことは容易だ。ソーシャルメディアには、既存メディア批判が溢れて

1 Trusted Web ホワイトペーパー Ver.1.0
https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/documents_210331-2.pdf

おり、誤報に対し謝罪した事例も検索すれば見つけ出すことができる。一方、ポータルサイトやニュースサイトは間違っても謝罪することはまれだ。

前述したように、ネットを流れる情報やニュースは、ソースや経路が不透明なため検証が困難であり、編集方針もアルゴリズムも提示されておらず、同じく検証することはできない。フェイクニュースや陰謀論に限らず、現状のネットメディアは内容真実性を確認することができない。とすれば、メディアリテラシーで批判的に読み解く対象となるのは必然的に既存メディアになる。メディアリテラシーを推進すれば、既存メディアを批判的に読み解き、信頼が低下する。そうすると、あらゆるメディアの信頼度が低下し、ディストピア的な世界が到来する危険性がある。

既に警告は行われており、ソーシャルメディアの研究者ダナ・ボイド（Danah Boyd）が、トランプ（Donald Trump）が米国で大統領になった直後に、メディアリテラシーだけで問題を解決することの危険性を指摘している[2]。この指摘は批判されたが、その後ますます分断が進み、米国においてもメディアリテラシーは十分な処方箋になっていない。

また、ニューヨークタイムズに「ウサギの穴に飛び込むな（Don't Go Down the Rabbit Hole Critical thinking, as we're taught to do it, isn't helping in the fight against misinformation.）」というオピニオン記事が掲載されている[3]。ウサギの穴とは「不思議の国のアリス」に登場する不思議の国への入り口のことだ。クリティカルシンキングのような批判的に物事を捉える方法はフェイクニュースに役立たず、むしろウサギの穴に飛び込むことになってしまう。つまり陰謀論にからめ捕られるということである。

フェイクニュースに対抗するにはメディアリテラシーが必要であると言う人は、こういう海外の議論や指摘をおおむね無視しているようである。注意してもらいたいのは、メディアリテラシーの重要性を否定しているわけではないということだ。情報の信頼性が多角的に検証できない状況で、メディアリテラシーを推進すればするほど、トラストを揺るがすと指摘しているのである。 これが、メディアリテラシーのパラドックスである。

ネットやソーシャルメディアは、便利だから使わざるを得ないが、そのトラストは不全であり、ネットやソーシャルメディアの信頼度向上という再構築を行うことなく、メディアリテラシーを推進することは、社会の分断や対立を増すだけである。

2.2.2 「ニュース生態系」と「信頼」の欠損

次に、インターネットメディアのトラストを再構築するための手がかりとして、ニュース生態系について述べる。

インターネット以前、ニュースは既存メディアだけが担っていたが、今やポータルサイトやニュースサイト、ソーシャルメディアとさまざまなメディアが担うようになっている。特に、ソーシャルメディアは、人々だけでなく、政治家や企業、既存メディア、Bot、プロパガンダと指摘されているアカウントなど、が発信する玉石混交の空間となっている。

ハーバード大学のヨハイ・ベンクラー（Yochai Benkler）らは、ソーシャルメディアだけでなく、テレビやラジオなどが有機的に結びついた情報の生態系が形作られ、フェイクニュースの拡散を助け、分断を煽る役割を担っていると指摘している。この、情報生態系が汚染されているという考えを基に、国内のフェイクニュース調査を行い『フェイクニュースの生態系』にまとめている。ソースや経路を追跡する調査で分かったのは、フェイクニュースはニュースが作られている生態系そのものから生まれているということである。

2 Danah Boyd, "Did Media Literacy Backfire? ", Data & Society,2017, https://points.datasociety.net/did-media-literacy-backfire-7418c084d88d

3 Charlie Warzel, "Don't Go Down the Rabbit Hole", New York Times, 2021, https://www.nytimes.com/2021/02/18/opinion/fake-news-media-attention.html

図2–2–1でフェイクニュースの生まれ方を一つ紹介する。テレビニュースをソースにツイートが行われ放送されていない情報が付け加えられる。それが、まとめサイトにまとめられる。まとめサイトの内容をニュースサイトが記事化し、ニュースとしてポータルサイトに配信され、さらにまとめサイトや掲示板、ツイッターなどに拡散していく。フェイクニュースは生態系におけるメディアの相互作用により生まれる。

ツイートやまとめサイトの段階では、拡散力は乏しい。ニュースサイトとポータルサイトの記事は同じ内容だが、ポータルサイトに配信された記事のほうが、まとめサイトや掲示板で拡散している。ここには、ネットのニュースにポータルサイトが強い影響を持つ日本のニュース生態系の特徴が現れている。

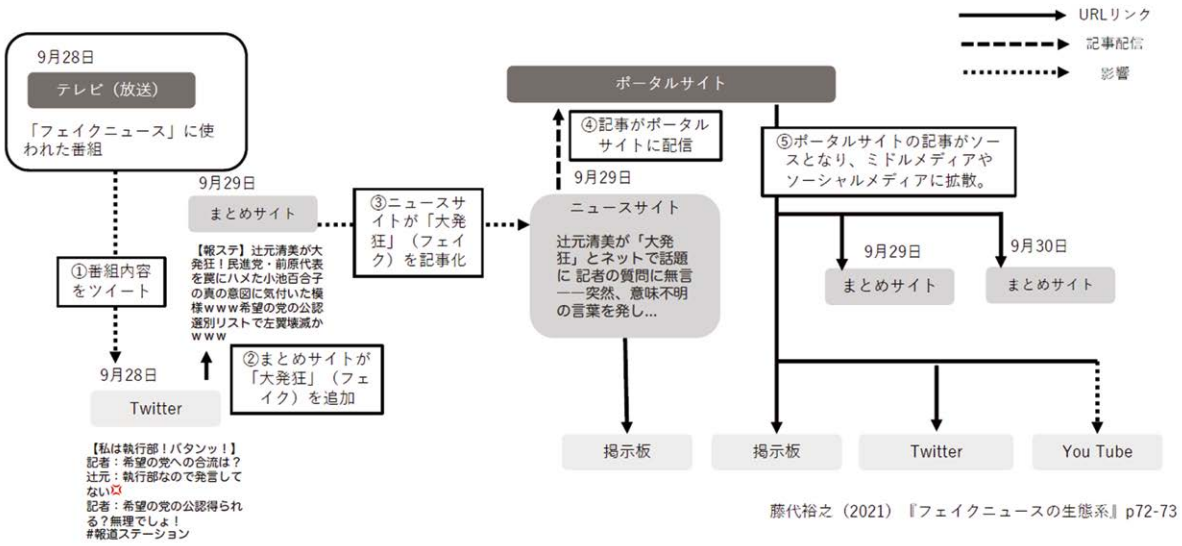


図2–2–1　「フェイクニュース」の生成・拡散過程

最近、ロシアのプロパガンダメディアと指摘されているニュースサイトを情報源にしたスポーツ紙の記事が、ポータルサイトに配信され、ソーシャルメディアで拡散する事態が起きた。これも図2–2–1と同じ構造で起きている。日本のニュース生態系は、外国のプロパガンダに対しても脆弱性があることが明らかになった。このようなことが積み重なっていくと、もともと乏しかったネットにおけるトラストがさらに厳しいものになっていくことになる。

フェイクニュース問題は、環境問題のようなものだ。ネットの汚染を改善するためには、ソーシャルメディアだけを規制しても意味がない。生態系そのものを把握したり確認したりしておかないと、どこかを止めたり、どこかだけを何かをするということにはならないだろうと考えている。

みずほリサーチ＆テクノロジーズによる「令和3年度 国内外における偽情報に関する意識調査：フェイクニュース及び新型コロナウイルス関係の情報流通調査結果」によると、ポータルサイトやまとめサイトではフェイクニュースに接することが多いという結果が出ている。他国に比べるとわが国においてはやや多いということである。一方、ポータルサイトとか検索サイトは比較的信頼できるという割合がネットメディアの中では高い。それに汚染が含まれているとなると、トラストを構築するにしても、八方塞がりの状況になってしまう。

さらに対策を難しくしているのが、ユーザーが自分の好みに応じた情報に囲まれる技術により生じる、フィルターバブルやエコーチェンバーだ。異なる人の間で共通の認識が成り立ち得ることがトラストの根源であるなら、フィルターバブルやエコーチェンバーの中にいる人たちにとって、ネットはトラストな状態となってしまう。

このようなニュース生態系の現状を確認した上で、トラストを再設計していく必要がある。

2.2.3 参考になる取り組み

トラストを再設計するために参考になる取り組みをいくつか紹介する。

図2–2–2に日本のニュースの生態系の構造を示す。ソーシャルメディア、ミドルメディア、マスメディアという3つの構造になっている中で、不確実情報がニュース化して汚染が広がる。この生態系それぞれのメディアのレイヤーに対して、あるいは、それを支える技術に対して何か対策するという方法がある。

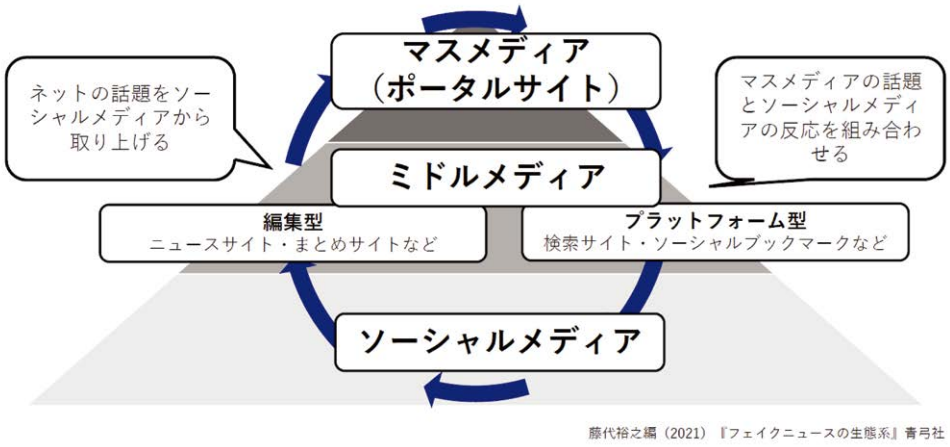


図2–2–2　　生態系の構造とミドルメディア

一つのポイントはミドルメディアにある。図2–2–1で明らかにしたように、フェイクニュース生成の中心となっているのが、ニュースサイトやまとめサイトといったミドルメディアである。現状ではポータルサイトのニュースの編集方針や検索サイトの表示順を決めるアルゴリズムなどは提示されていないが、フェイクやヘイトなど不確実な情報を発信しているミドルメディアをリスト化し、ポータルサイトや検索サイトに掲載しないという方法がある。また、不確実な情報を発信しているミドルメディアへの広告掲載を停止することで、活動を抑制するという方法もある。

米国では、ウェブサイトの信頼性を、ジャーナリストが複数要素から評価し、これをユーザー側のアプリに表示するという取り組みもある。リスト化の際に参考になるだろう。

次に、ニュース報道への汚染を防ぐことが考えられる。前述した、ロシアのプロパガンダメディアなど不確実なネットの話題を情報源にした記事のように、ニュースにフェイクやヘイトといった汚染が入り込み、蓄積されるという現象がある。影響力の大きいポータルサイトやテレビ放送に汚染をなるべく入れないようにするべきである。そのためにはトレーサビリティーの確保が重要になる。

トレーサビリティーというのは、記事を書く際に、不確実な情報を発信しているミドルメディアやソーシャルメディアを扱わないように、つまり汚染がニュースに混入しないようにする仕組みの構築のことだ。記事を書く際に情報源をチェックできる仕組みのようなものを想定している。ソーシャルメディアの話題をニュースに扱う点では、ベンチャー企業の取り組みもあるし、テレビ局もベンチャーのサービスを導入している。残念ながら現状では主に事件や事故を速報する方向に利用されており、信頼の向上に技術が使われているとは言い難い。同じような技術が、不確実な情報を発信するサイトのコンテンツを追跡するトレーサビリティーの確保やデータベースの構築に利用できるはずだ。

また、情報トリアージという方法もある。ネットが汚れていることを前提に、浄水器のようなフィルターを通すことにより、意思決定者に対して的確な情報を提供する方法だ。災害時のツイッターでの救助要請に対して

研究しているが（藤代ら 2018）、フェイクやヘイトにも応用できる仕組みと言える。機械学習や自然言語処理の知見が重要で、研究者らが精力的に取り組んでいる分野なので対策に応用できる可能性は高いと考えている。不確実情報を発信しているサイトのリスト、データベースが構築されると、より精度が向上するため、研究者とジャーナリストが連携して取り組む必要がある。

こういった研究を進めながら、インターネットメディアを私たちが確からしく使えるような状態を作っていくことが必要である。リテラシーが大事とか、メディアがフェイクニュースを流している、というような話になるが、その前に私たちが使っているネットの信頼性は低いという現実から出発しなければならない。その低い信頼性をどうやって向上させるかが、トラストという面からしても重要になる。

【質疑・討議】

稲谷：「コロナに関して「信頼」できるメディア」という調査において、日本が特殊な形になっている。これは日本の情報の生態系の特殊性によるものか？

藤代：国によって生態系はそれぞれである。

稲谷：メディアにある種の免許制を導入することによって、確実性の高いものを追跡可能な形で提示するというやり方もあると思う。これについてはどう考えるか？

藤代：免許制は表現の自由の観点から望ましくないが、おおむね信頼できるメディアとなるような自主的な取り組みはあり得る。これまで新聞やテレビは自主的に信頼を積み重ねる取り組みを行ってきた。番組審議会や第三者委員会などで自社の報道をチェックし、放送業界にはBPO（放送倫理・番組向上機構）もある。ネットメディアにはそれがない。編集方針も説明していない。まずは、第三者委員会が掲載するニュースや情報をチェックするという枠組みをポータルサイトやネットメディアに導入していくというのが、信頼を向上させる鍵になるだろう。

福島：出典など一次情報をたどりやすくする、あるいは、考えやすくするような道筋や仕組みはあるか？

藤代：引用先やソースをURLで示すというのは技術的に可能だ。トレーサビリティーのところでも説明したが、利用者よりも生態系の中の仕組みとして利用する方がよいと思われる。なぜかというと、スーパーで売られている野菜にもトレーサビリティーが示されていることがあるが、我々みんながトレーサブルな野菜を買っているかというと、そうでもない。メディアリテラシーが解にならないという現実を示している。トレーサビリティーを無視するメディアをポータルサイトや検索サイトに表示させない、もしくは表示順位を下げることが重要だ。

大屋：日本の場合、報道自身がデータや情報のソースを明示しない習慣に浸りきっていて、トレーサビリティー確保というトラスト確保の手段を毀損しているという問題がある。社会調査において信頼区間を示さない、政府の審議会等の報道で正式名称を示さない、科学報道で元論文のタイトルなどを示さないなどにそれが見られる。

藤代：ニュースは相互作用により生成・成長・拡大してくので、その経過を追えることや、問題あるものを拡大しないようにすることが重要だと考えている。

中川：米国の場合は、メディア自身が共和党と民主党に完全に分断されており、お互いにエコーチェンバーなどの状況になってしまい、それを解きほぐそうという努力はほとんど無駄であると言う人もいる。ただ、メインの発信源は民主党も共和党も3%くらいの人でしかないので、出てきたものが信用できないほど過激化してくれれば、みんなが信用しなくなるというストラテジーの方が、有効に作用するのではないだろうかと思われる。これに関してどう思われるか？

藤代：メディアリテラシーの重要性を訴えると、既存メディアを疑うようになり、過激化や分断が進む可能性がある。それよりは、一部のフェイクニュースやヘイト、過激な発言をしているサイトやアカウントの力をそいでいくようなトレーサビリティーやデータベースを作るといった技術とリテラシーを両面で展開する方がよいと思う。

2.3 振る舞い予想・対応可能性の視点から

杉村 領一

2.3.1 国際標準化におけるトラストの捉え方

国際標準化のための組織ISO/IEC JTC 1に、人工知能（AI）に関する分科委員会SC 42が2018年に発足し、日本国内にそのミラー委員会（SC 42専門委員会）も作られた[4]。私はSC42に参加しているとともに、国内のミラー委員会のまとめ役という立場にある。そこで、トラストに関して、AIの国際標準化の動向と絡めて話したい。

トラスト（Trust）の定義に関して、図2–3–1左の一つ目の定義が、本日冒頭の講演で示されたものとほぼ同じだと思うが、標準化などの国際的議論の場では、TrustやTrustworthinessなどの言葉の定義は種々ある。出版済みのISO/IEC 25000シリーズにおいて、図2–3–1左の二つ目のような定義となっている。二者・三者関係とされているが、AIが入ったときにどう考えるかなど、種々議論がある。

現在、ISO/IEC JTCのレベルでTrustworthinessのワーキンググループ（WG13）ができて議論中である。今後変更される可能性があり、また、版権上の制約もあるので、図2–3–1右にはその意訳のみを示す。基本的には、Trustworthinessは、実証可能、検証可能、測定可能な方法で利害関係者の期待に応える能力とされている。文脈やセクターなどにも依存するし、検証も必要だとされる（図2–3–1右の注1）。ただ、注2には、Trustworthinessの特性に、信頼性、可用性、可制御性、堅牢性、回復力、そして、説明責任や透明性といった社会技術（Socio Technical）的なものまで含まれているので、技術的にはまだ扱いにくい。

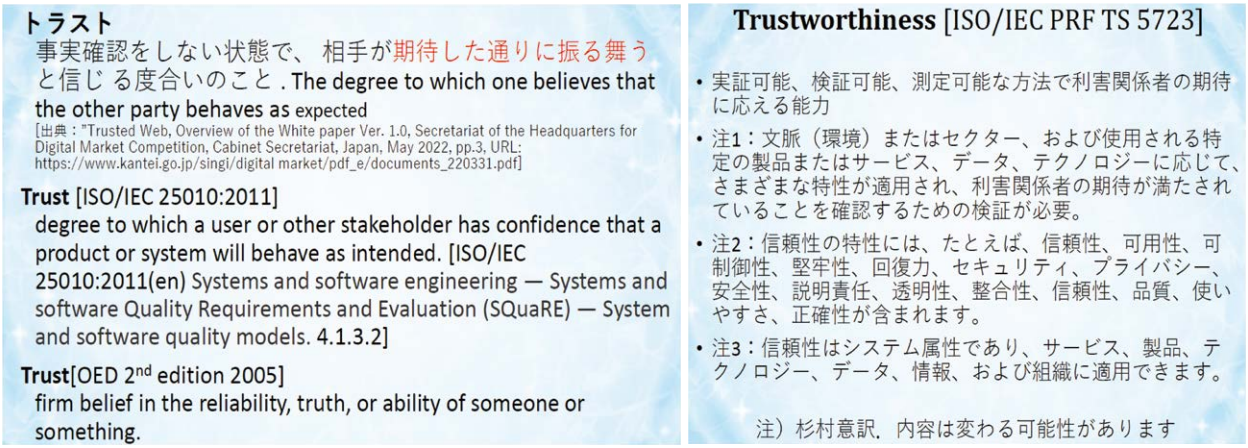


図2–3–1　　トラストの定義

4 ISOは国際標準化機構（International Organization for Standardization）、IECは国際電気標準会議（International Electrotechnical Commission）、JTC 1は第一合同技術委員会（Joint Technical Committee 1）である。国内のSC 42専門委員会については https://www.ipsj.or.jp/release/20180110_itscjnews.html を参照。

2.3.2 国際標準化の背景

標準化の背景として、AIについても話しておきたい。

まず、AIは人間の諸々の活動を近似する活動であるという捉え方がある。そして、その近似の精度が高くなってきたことで、逆にその近似から自分たち自身を見る鏡のような役割を果たすようになってきたと捉えることができる。ただ、世界には人間と機械を非常に強く峻別しようとする考え方もあり、AIが人間の領域を侵すと喧伝されて大騒ぎにもなった。その後、AIの利用方針を述べたドキュメントが、世界中から多数出た。その総数は600–700件以上と言われる。その中でも代表的なものを図2–3–2に挙げた。私たちも、全部を読めてはいないが、これらの内容を踏まえて、各機関が最終的にどのような形でAIに関する標準化をすべきかを議論している。

図2–3–2　　AI国際標準に影響を与える代表的文書（例）

2.3.3 日本の動きとISO/IEC JTC 1/SC 42

日本は世界の中では、かなり先行して種々のガイドラインが発表されており、例えば、経産省からAI契約ガイドラインやAIガバナンスのガイドライン、産総研から機械学習品質マネジメントガイドラインが出され、経団連からは欧州AI規制法（AI Act）案に対するコメントも出されている。 SC 42国内専門委員会としては、これらをベースに方向性を定めて取り組んでおり、基本的にはソフトロー指向、リスクベースのアプローチを採っている。欧州がAIに関するリスクを多段に分けているが、ハイリスクAIへの厳しすぎる制約や不明確な基準が産業育成を阻む恐れがあり、判断基準の明確化等の要請が日本から欧州側に出されている。

また、欧米からはAIに関するリスクとしてネガティブなものばかり挙がってくるが、本来、期待からのずれとしてのリスクにはネガティブとポジティブの両面がある。日本からは、AIのポジティブな面も考えるべきだという意見を出して、ガバナンスのガイドライン等に反映させるなど、AI特有の特性から来るリスクの側面について標準化に反映させてきている。

SC 42の状況は（図2–3–3）、Trustworthinessを含めて非常に広範囲の議論があり、ワーキンググループだけで7つ設けられている。35か国がパーティシパントメンバー、15か国がオブザービングメンバー、計50か国が参加している。総会も2018年から9回開催されてきた。2022年4月の第9回総会には約250名の参加者登録があった。国内組織であるSC 42専門委員会には、現在30組織が参加している。

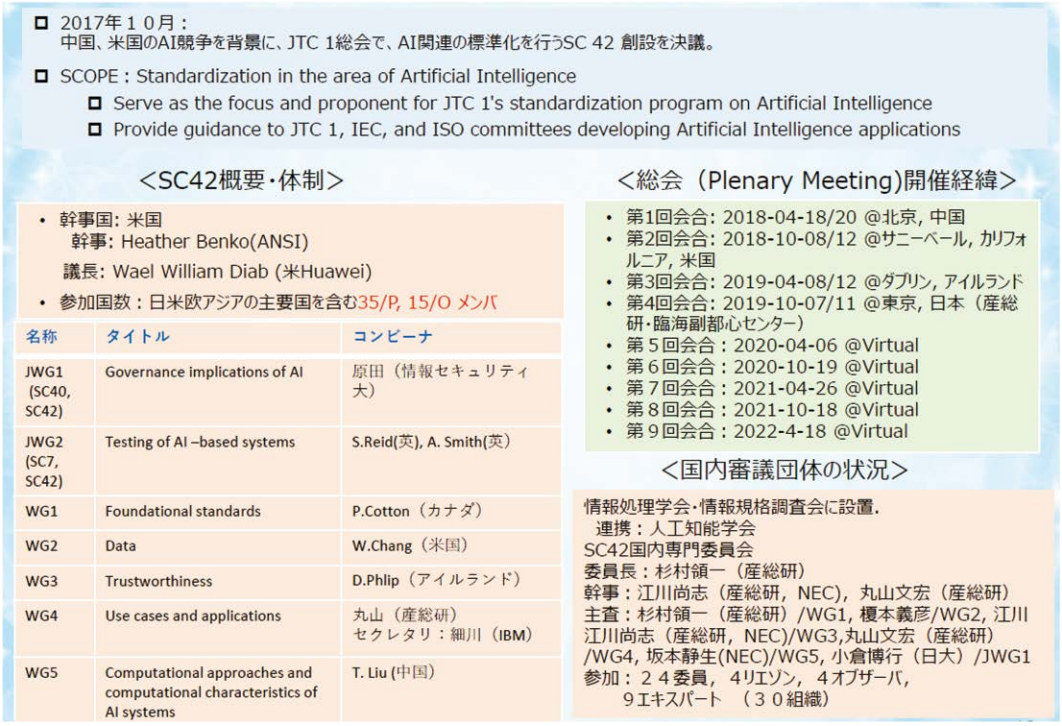


図2–3–3　　SC 42の概要

このような標準化活動において、日本側の意見をうまく反映していけるよう、コンビナーやエディターを輩出している。現在、全体で37のワークアイテムが進んでいるうち、11について日本が大きく貢献できている。ただ今後も難しい議論が続いていくため、国際的なプレゼンスを保ちながら、日本の産業界の阻害要因になるような標準策定は避けるよう、活動を進めている。

2.3.4 AI国際標準化におけるTrustworthiness

Trustworthinessについて考えるには、ガバナンス、テスト、データ、用語の基本、ユースケース、計算論的アプローチなど、非常に広範囲な議論をせざるを得ない。幸い、これらは全部、ワーキンググループの形で議論が進められている。国際標準は高い抽象度を保ち議論する側面があるため、本当に対象が捉えられているのかという視点でも見る必要がある。

関連するワークアイテムとして最初に出版されたテクニカルレポートが「Information technology – Artificial intelligence（AI）– Overview of trustworthiness in artificial intelligence」（ISO 24028:2019）で、AIに関するTrustworthinessのオーバービューをまとめている。まずは、最新の知見を集めて出版し、フィードバックを得ることを目論んだが、JIS化の予定もないためか、フィードバックはあまりない。

ガバナンスについては、日本の情報セキュリティ大学院大学名誉教授の原田要之助先生が国際コンビナー

となって、特にガバナンスボードの役割などについてまとめていただいた「Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations」（ISO/IEC 38507）が今年出版された。今年JIS化も予定されている。また、プロセスの標準については、「Information technology – Artificial intelligence – Management system」（ISO/IEC 42001）の作成が進められている。

今回の主題テーマであるTrustworthinessという概念そのものは、AIだけでなくIT分野全体で非常に重要なものなので、現在、JTC 1のWG 13で「Trustworthiness vocabulary」（ISO-IEC TS 5723）の作業が進められているところである。

2.3.5 関連する国際的な動き

欧州はリスクベースのアプローチを採り、AI規制法案ではリスクが「許容できないリスク」「高リスク」など4レベルに分類されている[5]。これに関連する標準については、JTC 21で議論が行われている。あと2年ほどで法制度に落とし込む計画になっており、日本も非常に高い関心を持っている。標準化関連のメンバーはオブザーバーとして入る予定で動いている。AI Watchが整合規格（Harmonised Standards）として挙げた例[6]が全部で12個あるが、そのうち6個を日本がリードしている。

一方、これと全く異なるコンテクストだが、米国のNIST（National Institute of Standards and Technology）がAI Risk Management Framework（AI RMF）を出した。この中では、用語をTechnical、Socio Technical、Guiding Principlesという3つに分けて（図2–3–4）、それぞれについて議論をしっかり行おうとしている。

図2–3–4　Socio-technical perspective in AI RMF by NIST

さらに、関連用語として、バイアス（Bias）、公平性（Fairness）、アカウンタビリティー（Accountability）、説明可能性（Explainability）、監査可能性（Auditability）等々あるが、AI標準の中では、これらの用語をできるだけきちんと深掘りしようとしている。ただ、昨今のAIの開発速度と比べると、議論にはかなり時間

5 https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

6 “AI Watch: AI Standardisation Landscape state of play and link to the EC proposal for an AI regulatory framework” (2021) の図18
https://op.europa.eu/en/publication-detail/-/publication/36c46b8e-e518-11eb-a1a5-01aa75ed71a1/language-en

を要している。

いずれにせよ、AI標準関連の動き全体の地図を描く動きはようやく出てきている。SC 42でもロードマップを作っている。OECD（Organisation for Economic Co-operation and Development：経済協力開発機構）や欧州などがObservatoryを設置し、世界中の文献の位置付けなどを分析しやすくする環境を作っている。また、これだけ新しい概念や標準が出てくると、全体の関係性と既存の標準との関係性が分からなくなるので、エコシステムアプローチという考え方（図2–3–5）も提案されている。

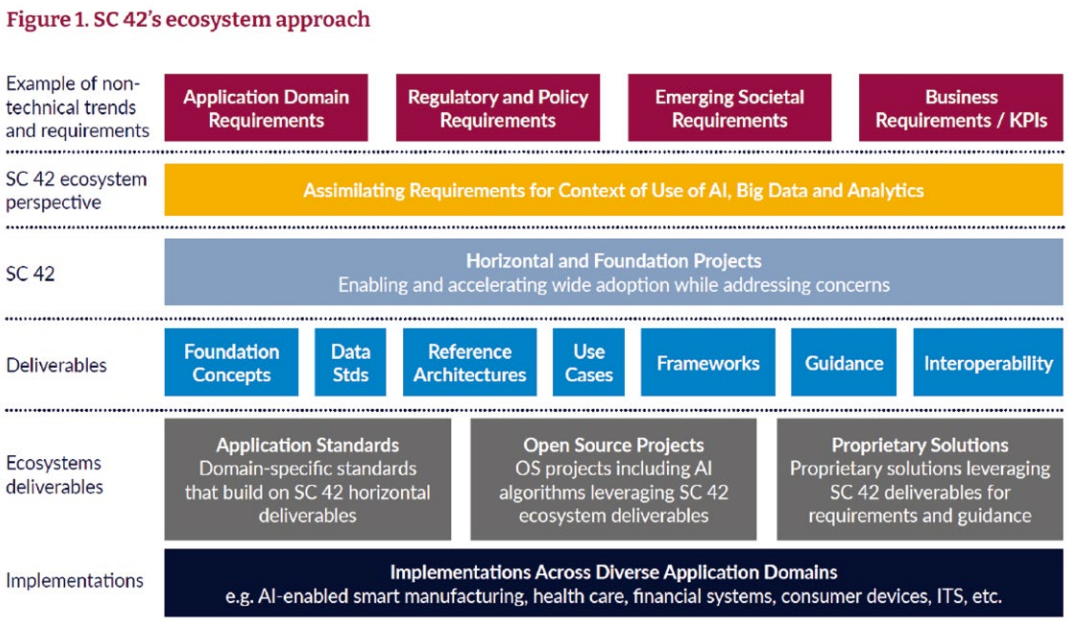


図2–3–5　Ecosystem Approach

2.3.6 今後の課題

今後必要になってくるものとして認証制度がある。今年（2022年）になってから世界中で動きが出ている。

認証制度については、WTO（World Trade Organization：世界貿易機関）によるTBT協定（Agreement on Technical Barriers to Trade）というものあり、国際標準が成立した際に、同様の国内規格が存在する場合には、国際標準の方を優先することが義務付けられている。このTBT協定があるため、認証を実施する場合、国際標準側にしっかりと国内の利害も反映させて本当に使えるものにしておく必要がある。

もう一つ、昨今、研究開発の対外発表において、ソースコードとデータの提出が条件になってきているが、同様に、認証の際に、ソースコード、データ、そしてそれらのライセンスをどのように扱うかが課題になるだろう。AI標準化は、研究開発、認証、特許、ソースコード、データ、ライセンスなど、広く見て取り組まねばならないものになっている。

今後の取り組みを考えていくと、我々は人間と機械の関係を改めて考えていく必要があるだろう。監視（Oversight）という言葉も最近よく使われる。Human over the loop、Human in the loopなどの言葉もあるが、まだ種々の定義があり、今一度、人間と機械の関係という観点から種々の側面を深掘りする必要があるだろうと考えている。私の個人的な見解としては、部分情報しか持てない状況（Information Partiality）や不確実性（Uncertainty）などに基づいて、議論を整理する必要があるのではないかと思っている。

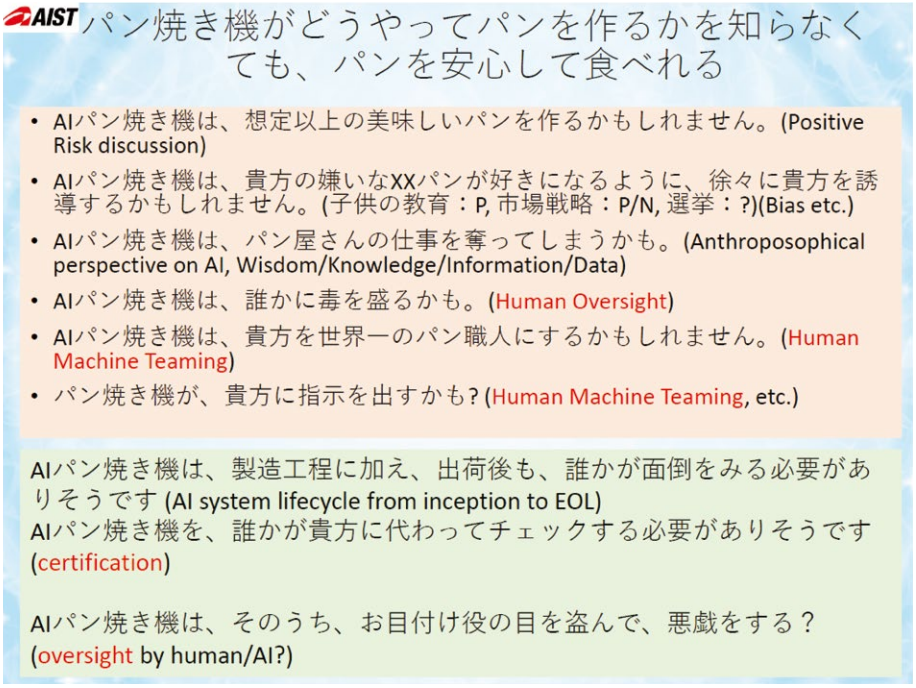


図 2-3-6　　人間と機械の関係のさまざまな形

【質疑・討議】

福島：AIはブラックボックスと言われ、確率的・帰納的で動作保証ができない。説明可能AIの技術もあるが、結局、近似的説明でしかないので、100%保証はできない。これらの現状から、社会がAIをトラストして受け入れるようになるのは、どのような方向性だと思われるか？　例えば、AI開発のプロセスをしっかりやったことを認証してもらうとか、事前に100%保証はできないので、事後評価というか、問題が起こったところで責任や保険がしっかりしているとか、経験を積み上げてトラストを作っていくとか。

杉村：いま例示されたような方向性で議論されている。ただ、プロセスについては、従来は出荷までのプロセスしか見ていなかったと思うが、その後まで含めたライフサイクル全体を捉えて、監視・監督していく必要がある。これについては、日本からライフサイクルに関するドキュメントを提案し、幸いこれがアクセプトされ、現在作成中である。実はこの議論も大変だった。最初は出荷後を考えておらず、エッジでの学習など市場に出てからの学習のもたらす新たな側面はケアされていなかったが、それでは駄目だという議論になった。さらに、問題を起こしたときにどういう対処をすべきかについては、かなりシビアな議論になっている。まずはテクニカルレポートのレベルで課題形成を進めている。

福島：事後評価に任せるだけでは駄目で、やはり事前の保証がかなりしっかりしていないと、社会に受け入れられないという立場の人が多いということか？

杉村：そのように感じる。 Verification、Validation、Testingなど、どのプロセスでどこまでやればよいか、かなりシビアな立場の意見もあり、全部仕切り直しになるような議論も行われてきた。

茂木：欧州が人間と機械を厳しく峻別するという話だが、そのような方向で国際標準が決められると、フレンドリーなAI・ロボットを目指す研究開発に支障をきたす可能性はないか？　AIの良い面を生かせなくなるなどの不都合が起きないかという危惧がある。

杉村：その通りで、我々も非常に苦労している。まず、欧州の人たち全てではないが、多くの人々は擬人化表現を蛇蝎のごとく嫌う。そのため、AIの定義自体が大議論になって決められなかった。結局、AIシステムという形でなんとか妥協してもらった。もう一つは、ウィズダム、ナレッジ、インフォメーション、データという階層はよく知られているが、欧州の人たちから、ウィズダムとナレッジは人間しか持たな

いもので、エンジニアリングで扱えるものではないから、用語集から「ナレッジ」を全て削除しろという主張があった。 AIには、Knowledge Representation等の用語があるにも関わらず、このような主張があり大議論だったが、最終的に差し戻すことができた。欧州の価値観・コンテクストから強くセンシティブな主張をしてくることがあり、欧州を中心とした動きに対して、日本は中立的なスタンスを取りつつ、他の国と協力して動いている。

茂木：米国はどのようなスタンスか?

杉村：米国は一つの意見ではない。

福島：AIのトラストを確保するために、AIは道具だという側面できっちり押さえ込むという方向がある一方、自律性の高いAIと人間との間のトラストをどう考えるかというのも、非常に興味深い研究課題になる。しかし、欧州的なセンスからすると、前者のみで、後者は完全に排除されているということか?

杉村：そのように感じているが、我々としては、Oversightの考え方を入れて、AIと人間との関与の仕方をきちんと技術的観点から再考して整理しないかと提案している。本来、可能性としては議論しておくべきと思う内容に関して、かなり感情的な反応が返ってくることがあり、苦労している。

2.4 法制度・ガバナンスの視点から

稲谷 龍彦

Society 5.0における法とガバナンスという観点から話題提供する。この報告において信頼（トラスト）とは「相手が思い通りに行動しない可能性が残存していてもなお、相手が自分のために行動すると考える心理状態」という山岸俊男先生の定義に近いものを採用する。信頼の存在は情報費用を低下させるため、未知の人との交流や取引などの関係性の構築を頻繁に生じさせることになる。日本では、未知の第三者に対する信頼性が相対的に低いと言われているが、グローバル化の進展やSociety 5.0の実現によって、未知の人やAIシステムのような事物との交流が盛んになると、低い信頼が高い機会費用を生じる可能性がある。つまり、信用できないからこの人とは付き合いません、こんなものは使いませんということを繰り返していくと、それらの機会を使うことで得られた利益が失われていき、かえって高い費用になる。

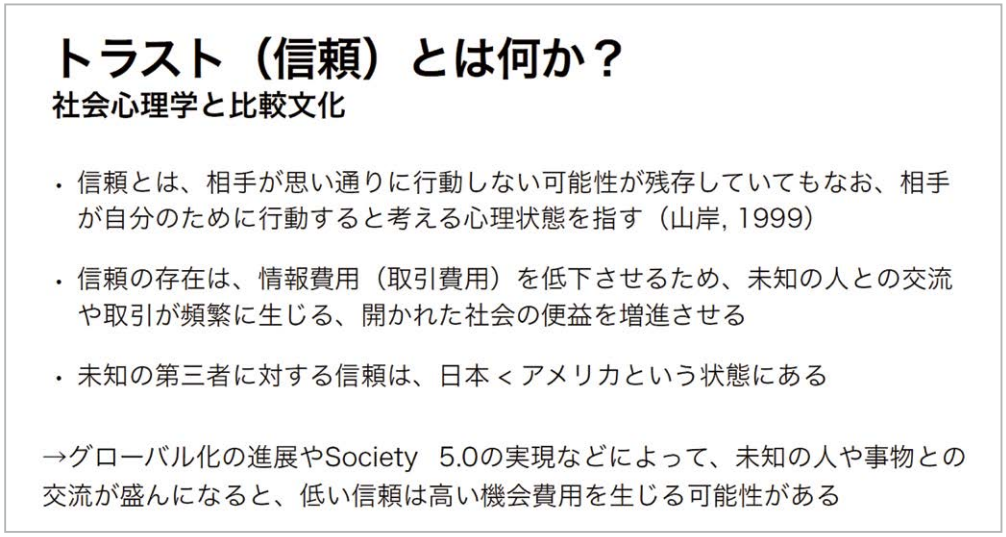


図2–4–1　　トラスト（信頼）とは何か？

文化人類学で言われるように、人間や事物は真空状態に存在しているわけではなく、社会システムのさまざまな構成要素と相互作用しながら存在している。文化や法制度といった構成要素は、人や事物の理解を生むフレームワーク、つまり、どのように人々や事物を認識するかに関係するため、それらに対する信頼の醸成にも影響する。文化差の話をすれば、AIと聞いて鉄腕アトムを想定するのか、それともターミネーターを想定するのかも、まさに構成要素が影響していると言えるだろう。

法は人々が世界を見る見方を内面化するものであり、その見方の内容にも、内面化を進めていく過程にも影響する。法は人為的に作ることができるため、内面化の規範や、その過程について操作可能な変数と考えることができる。したがって、法の設計を通して社会の在り方、ひいては、人の在り方や事物の存在の仕方を再設計できる可能性がある。このような発想こそが社会工学としての法であり、法やガバナンスの設計を通してどのように信頼を獲得するかという議論ができる。

以上の観点から見たとき、現状の法制度は、信頼を構築する社会工学としての法という方向にはあまり向かっていなかったと言える。例えば、モラルハザードに対抗する古典的方法というのは、閉じられた関係性を構築し、限られた相手と繰り返し取引するネットワークを作り、その中で何かおかしなことをした者をネットワークから排除するという方法である。

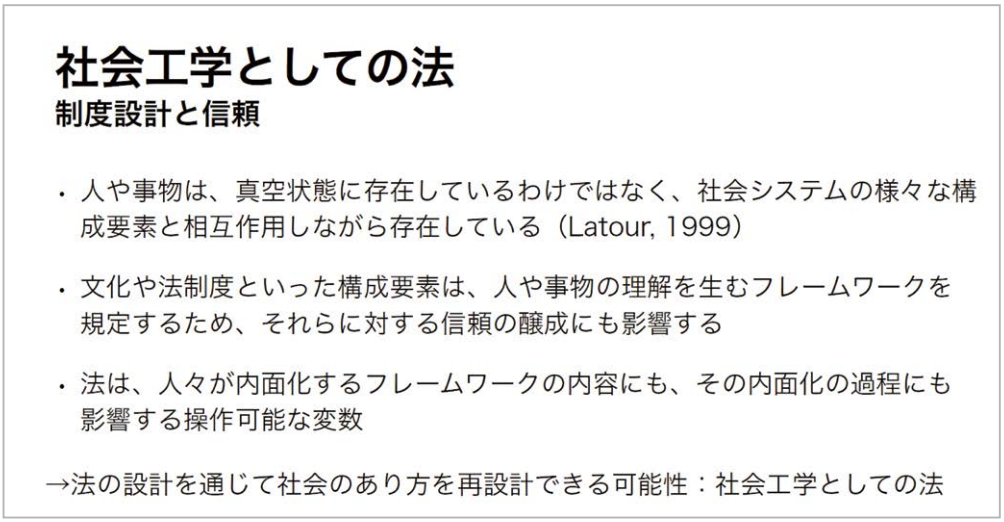


図 2–4–2　　社会工学としての法

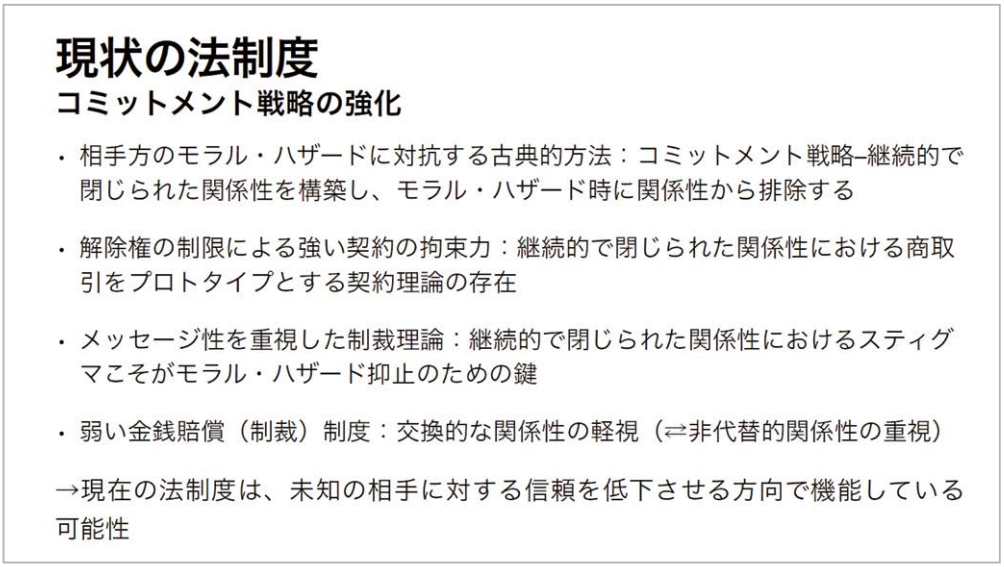


図 2–4–3　　現状の法制度

日本の法制度は、むしろこの閉鎖ネットワーク戦略を強化する方向でこれまで進んできたと解釈できる部分が多い。例えば、日本の場合、契約解除権の制限条項が契約書に書かれていなくとも、裁判所の法解釈によって制限が課されることがある。これは、よく知った人同士が契約関係に入ることが暗に前提とされているからであり、信頼関係が破壊されるまでは契約が解除できないと表現されている。逆に言うと、契約関係を容易には解除できないため、相手方をよく調べる必要が生じ、知らない相手とはなかなか契約関係を結べないことになる。このように法制度が、社会の在り方と共進化するような形で、今の社会の在り方を規定している。

これと歩調を合わせるように、日本は金銭賠償制度で十分な金銭を回収するのがしばしば困難である。その理由は、日本が特定の関係性、なんらかの交換不可能な特別な関係性というものを重視する傾向がある社会であるからかもしれない。同様に、閉鎖されたネットワークで大きな意味を持つ、メッセージ性を重視した制裁が社会的に重視されてきたと言える。

現在の法制度というのは、閉鎖されたネットワークへのコミットメントを進める方向に向かっており、その反面、知らない人に対する信頼を低下させる方向で機能している。したがって、現状の法制度を維持すること

を前提とした場合、未知の人や事物への信頼の醸成にはなかなかつながらない可能性がある。

現状の法制度の在り方をいきなり変えるのは難しい。法を含む制度というものは、人々が世界をどう見ているかという文化との関係で決まる。また、既に存在している人々の信念と大幅に異なる法制度やルールを施行しても無視される傾向があると言われている。以上のことを踏まえて考えてみると、現状の日本の法とは、欧州で発展してきた独立した個人同士による、特に言語を通じた関係性構築を重視する文化圏で誕生した近代法というものを、関係性を重んじる人々の、情動的理解を通じた関係性構築を重視する文化圏において変容してきた結果として、先に説明したような特殊な状況になっていると言えそうである。

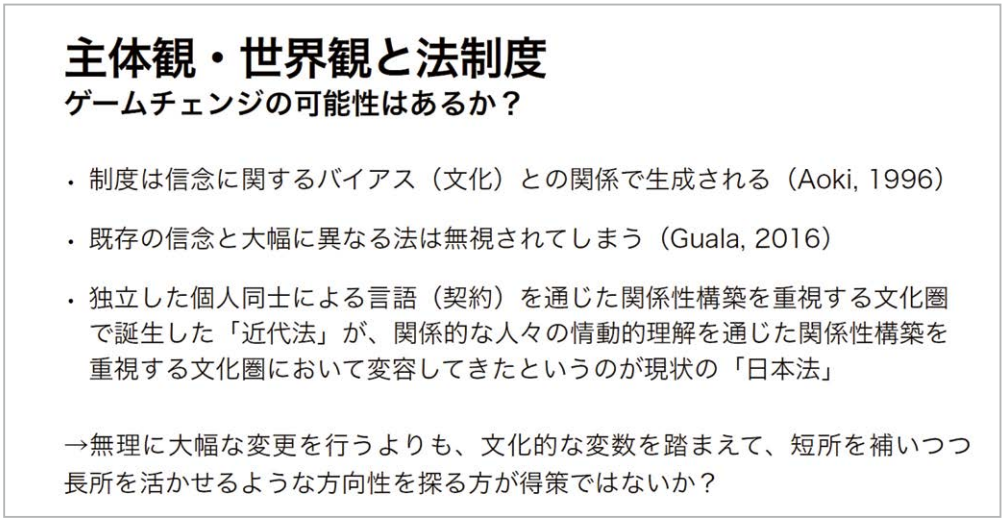


図2–4–4　　主体観・世界観と法制度

ただ、そうだとすると、いきなり未知の第三者に対する信頼をベースにするような法を導入できたとしてもうまくいかない可能性が高い。文化的な変数の存在を踏まえた上で、この変数の短所を補いつつ、長所を生かすような方向性を探る方が得策かもしれない。

他方で、人工知能を活用した自律的なロボットは、言語や理性を介して人だけが本来の関係性を構築できると考える西洋近代法の世界観とは相性が悪い。つまり、自律的な人工物を活用していく社会像と合っていない可能性がある。実際、倫理的なAIシステムに関する要求事項では、説明可能性や支配可能性が重視されており、理性や言語による身体・事物への支配が優位な文化圏のバイアスが顕著に現れている。

それに対し、日本人は、言語ではよく分からないが、何となくうまくいったり、何となく馴染んだりするものを受け入れることができる点に長所があるとすると、実は日本だからこそ追求できるAIシステムの可能性があり、それが何かしらの大きなゲームチェンジの可能性を持っているかもしれない。

仮にそうだとすると、人とAIシステムを馴染ませるためにAIシステムと共に変化し続けていくための制度を準備する方が良いと考えることもできる。これと関係するのがアジャイルガバナンスの考え方であり、そこで提示されるような人間観や世界観を掘り下げて、世界に発信していくという戦略も一つの可能性としてあるだろう。

賠償制度や制裁制度は事後的な評価に重点があるが、人々とシステムをより良く馴染ませるという観点から考えたとき、例えば、事故が起きたら事故の原因の究明とかシステムの改善に重点を置き、なるべく情報の共有や問題解決のために人々の知恵を集合させることを推奨し、それを積み重ねていき、時間をかけて人、制度、AIシステムの三者が馴染みながら、良い方向に変わっていくという戦略が考えらえる。

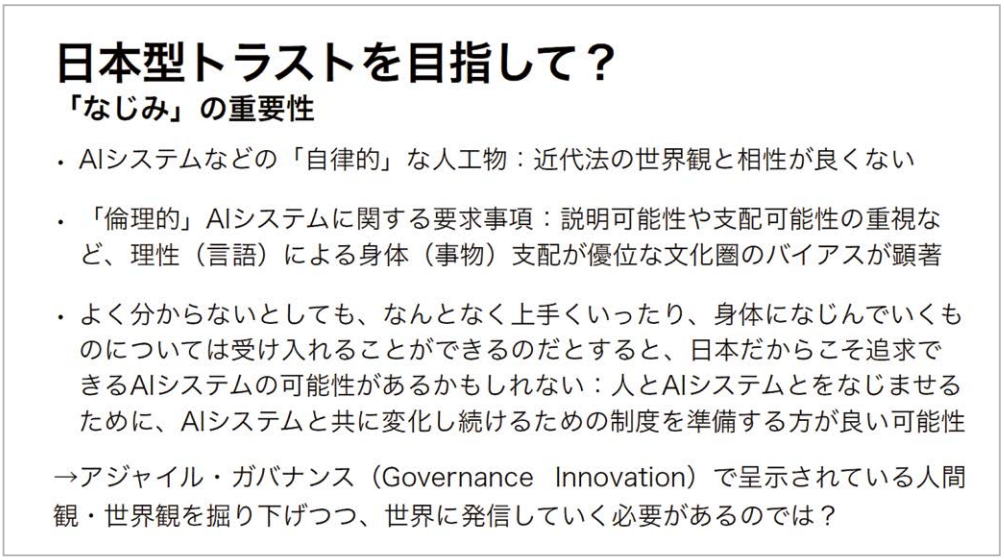


図 2–4–5　　日本型トラストを目指して？

もっとも、より高次で抽象的な価値としての幸福の実現のように、人々の多様な価値観をうまく集約しなければならない場面では、マルチステークホルダーアプローチのような、より機能する民主主義のシステムを作っていくことも重要である。いずれにしろ重要なポイントは、人と制度、そしてAIシステムが共進化していくことであり、それをガバナンスを工夫することで進めていくという点である。国内では、例えば、デジタル庁でアジャイルガバナンス原則が採択され、具体的な政策レベルまで議論を進めている部分もあるが、ガバナンスシステム全体を見渡したときには、AIシステムとの関係について、人およびシステムをどのように変化させていけばよいのかに関する提案はまだ少ない。国際的に見ると、日本の議論は先行している部分も多く、ポジティブな反応をされることも少なくない。このコンセプトをうまく国内で共有し、世界に対して積極的に発信していくことは、社会システムを実現するパッケージをより広いステークホルダーと共有して進めていくという方向性につながるだろう。

【質疑・討議】

大屋：基本的に同意だが、日本における農村共同体は高度経済成長に伴う人口移動で崩壊しており、その時点ではまだつながっていた「帰省」によるつながりも子世代への移行で断ち切られ、その代替として登場した会社共同体もバブル崩壊以降の産業構造変化によりもはや存在しなくなっている。そのような環境で、なお我々は「関係的」なのか。また、グローバライゼーションや外国人の人口増などにより、我々が独自の世界を維持できる／する可能性は失われてきているのではないか？

稲谷：少なくとも、独立した個人という主体を確固として持っている人は、それほど多くないのではないかという印象を持っている。刑罰のメッセージ性に対する執着などを見ていても、やはりつながる世界観は良くも悪くも維持されているのではないか。その上で、つながる人々や事物が多様化していく未来において、どのようなガバナンスを想定することができるのか。そこが、どのようにマルチステークホルダーアプローチを実装するべきなのかと関係する重要なポイントだと思う。

大屋：別の言い方をすると、自律的な個人というのは欧米社会においても近代法システムを支える理念型にすぎず、現実にはほとんど実在しなかったと言えるだろうし、情報環境の複雑化によって自己決定能力が相対的に低下した現代においてはなおさらそうだということになる。であればむしろ、日本社会に適した新たなガバナンス手法というより、現代的環境に適したものという切り口になるのではないか。その先端的情況が日本にある。

稲谷：近代的個人に関して、濃淡はあるが、その通りかなと思う。私としても、アジャイルガバナンスなどについて、そのような見せ方も十分あり得ると思う。ただ、その見せ方をすると、非常に嫌がる方も少なくないので難しいところだと思う。

中川：EUのAI規制法案のアンアクセタブルAIに関する部分では民主主義的な価値観が書き込まれており、対中国政策をはっきりと打ち出しているところがある。それはそれとして受け止めるとして、もう少しローリスクになると技術を入れるにしろ、制度を入れるにしろ、日本として工夫の余地があるのではないか？　アジャイルに対処する部分と、原則重視で行った方が良い部分についてどう考えるか？

稲谷：ハードローは時間をかけて多くのステークホルダーが納得する形で作っていくものであり変化するのに時間がかかる一方、ソフトローや個々の企業の工夫を活用するような世界になってくると、回転の速度と、その回転によって決定する事柄が、誰にどのぐらい影響を与えてくるかという変数を考えていくことになる。 EUとも合意できるところは合意しながらうまく進めていくのが賢いやり方だと思う。向こう側に乗っかってしまうと使えなくなってしまう可能性の高いもの（例：高度な医療AI）もあり、そこは腕の見せどころ。

村山：世界を見渡したとき、私たちが「日本的」といっているものが、世界の別の地域でも同じことを考えているかもしれない。世界で共通して押さえておかなければならないものについてアジャイルガバナンスで達成できるか、それとも文化圏、人種、あるいは国家でカテゴライズされる部分は別に進めていかなければならないか？

稲谷：日本的とされているものが、他の国・地域にも存在することはある。アジャイルガバナンスでカバーしており、またEUも気にしているポイントは、過度な権威主義やテクノクラシーのような、専門知が人々に説明責任も果たさずに統治する状態にはならないようにすること。欧州では、一個一個の機会、契約にかなり高い説明可能性を求めているが、我々はシステムとしてそこをどう担保するかに関心があり、そこは違う部分。

村山：米国はどうか？

稲谷：AIに関して言うと、いきなり規制をかけるのではなく、実験的にやっていくことをある程度重視していると理解している。ただその際に、例えば、人種差別が起きないようにするなど、かなり配慮して規制を作ろうとしている。

3 ｜ 社会・産業および人文・社会科学の視点から

第１章・第２章の発表を受けて、３名のコメンテーターからコメントをいただいた。各コメントから派生した質疑・議論は第４章の総合討議に含めた。

3.1 産業界の視点から

浦川 伸一

私は日本IBMを経て、2013年から損保ジャパンに勤めている。今、デジタル変革によって、ビジネスモデルが垂直統合（縦割りの産業）から水平統合（機能別に企業間で接続）へと劇的に変わるということが欧米を中心に進み出していて、日本にもこの波が来ている（図3–1）。企業間の連携についてはそれぞれのシステムやデータに信頼関係がないと成り立たない。

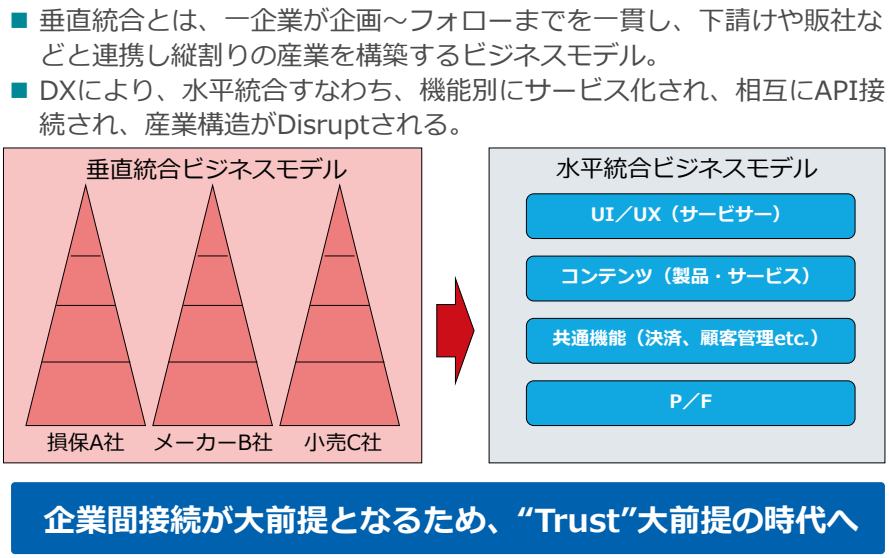


図3–1　垂直統合から水平統合へ

トラストの「３つの側面」は、産業界でも重視すべき観点である。対象真正性については個人／法人の認証に関するトラスト、内容真実性については保管データ・情報のやり取りにおけるトラスト、そして振る舞い予想・対応可能性については人工知能による処理に対するトラスト（説明可能なAIなど）が関わる。

産業界が特に注目しているトラストには、３つの分野がある。一つには、インターネット技術に関して、内閣のデジタル市場競争本部が主導するTrusted Web推進協議会があり、私も委員を務めている。インターネット技術は、デファクトスタンダードが主導して、市場で勝った人がマーケットを形作り、そこで標準が生み出されていくことが当たり前で、その結果、メガプラットフォーマーによる寡占やプライバシーへの懸念が大きくなって、慌てて規制を作り始めている状況である。

2つ目に、人工知能におけるTrusted AIは、すでに幅広く議論が広がっている領域である。また、3つ目に、今苦労しているのはサイバーセキュリティーであり、これはゼロトラストを前提に防御を定めなければならないということがうたわれている。企業を始めあらゆる事業体が莫大なコストを負担するという、極めて皮肉な状況に陥っている。

今日は1点目のインターネット技術にだけ触れる。インターネット技術は、「データ交換」の時代から、「情報閲覧／Webアプリ」、「データ共有・配信」、「連携型ビジネス」と便利になっていったがゆえに、使い方は年を追うごとに連携され複雑化していった。これらの4つの技術では、順に「通信の信頼性」、「サイトの情報の信頼性」、「データの信頼性・公平性」、「これらすべての信頼性」が必要である。

こういった背景から、内閣のデジタル市場競争本部でここ2年ほど、Trusted Webについて検討している。「監視社会」でも「一握りの巨大企業への依存」でもない第三の道を目指し、持つべき4つの機能を挙げて、実証実験などを行っていく（図3–2）。

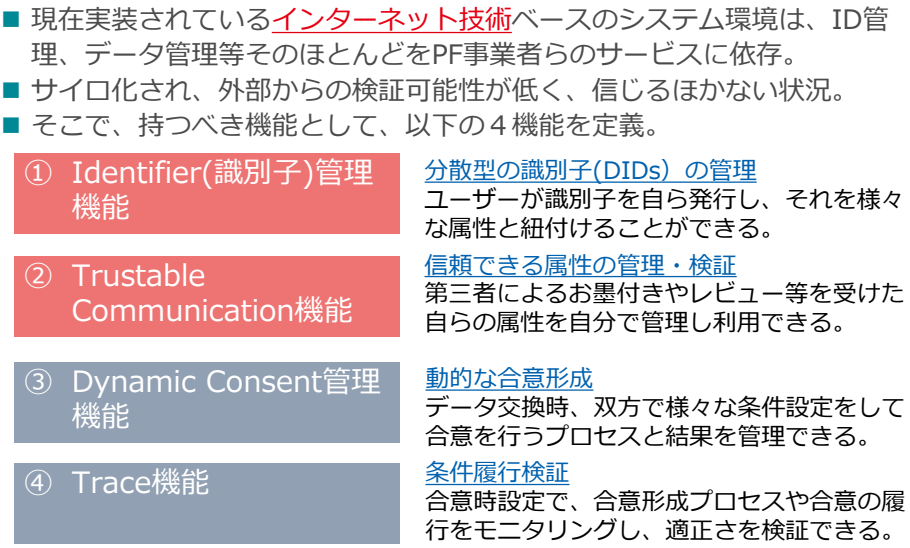


図3–2　　Trusted Webを司る4つの機能

産業界から見たときに、Trusted Web実装に向けた4つの論点として、信頼の定義と範囲、デジタルトランスフォーメーションとの関連、データ共有・個人情報保護への配慮、そして、ガバナンスの実装方法が挙げられる。ガバナンスはテクノロジーだけでは無理であり、政府や標準団体も関与するさまざまな形でのマーケットの形成が必要である。

■ 産業界としてこの構想を実現するにあたり、考慮すべき点を以下の通り四つの論点に整理してみた。

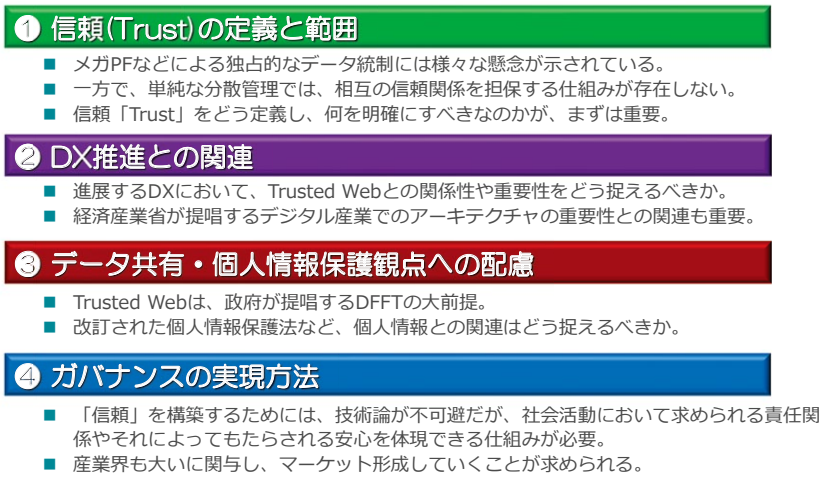


図 3-3　　Trusted Web 実装に向けた四つの論点

3.2 人文・社会科学の視点から（1）

川名 晋史

私は国際政治学の研究者なので、社会科学全体を代表するような議論はできないが、政治学に限って言えば、トラストという概念は古典的かつ重要なテーマである。しかし、これはまだ現実には見たことがなく、遠くにあるものだと考えている。政治学者にとってトラストを考えることとは、厳然たるディストラスト、不信の問題を考えることであり、もっと言えば、相互不信のメカニズムを探ることである。

そのことを踏まえて、自律型の人工物が現実的にはまだ先の話だとすれば、差し当たり政治学にとって重要になってくるのは、人工物の「後ろ」にあるもの、つまり、それを操作したりプログラムを書いたりする人間や組織、あるいは国という枠組みと、それを使用する側との間にある不信の問題をどう考えるかである。プログラムを書く人、組織、国の動機を可視化しないことには、恐らくこの不信というものを拭うことはできないだろう。また、特定のサイトや商品、行動、思想への誘導を技術的にどう防ぐかについても考えなければいけない。

これをパラフレーズすると、今回、対象真正性や内容真実性の保証が重要であることに同意する一方、それに加えて、「見張りを見張る人をどう見張るか」ということが第４のトラスト領域として設定される必要があるだろう。見張りというのは、技術の問題で、どういう形で対象真正性や内容真実性を伴うような技術を開発するかという問題である。しかし、その後ろ側には人間や組織がいる。それが「見張りを見張る人」だが、例えばそれが国や行政といった場合でも、それをどう信用するのかという問題が残るので、それをどう見張るかという、まさにこれまで政治学あるいは社会科学が取り組んできた問題に結局のところ立ち返ることになる。

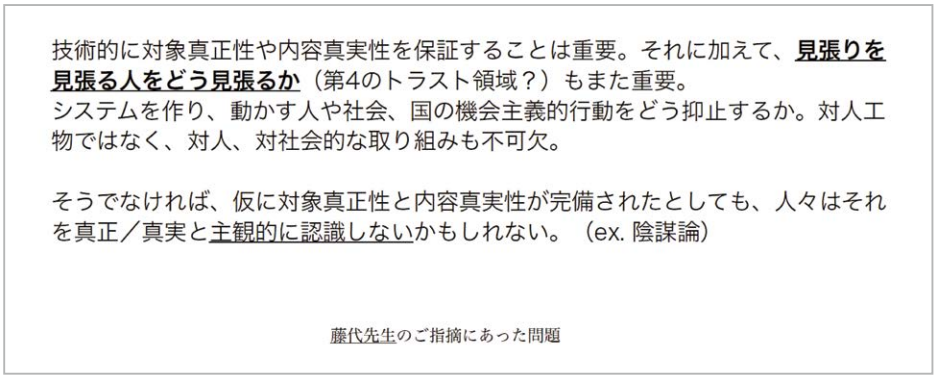


図3–4　　見張りを見張る人をどう見張るか

言い換えると、システムを作り、動かす人や社会、国の裏切りや機会主義的な行動をどう抑止するかが大事な問題になり、対人工物ではなく、対人、対社会的な取り組みも不可欠である。そうでなければ、仮に対象真正性と内容真実性が完備されたとしても、陰謀論に見られるように、人はそれを真正／真実であると主観的には認識しないかもしれない。これは藤代先生の議論と接続してくる。

これらを踏まえて、デジタル時代のトラストに私なりにどうアプローチするか。社会科学のオーソドックスな議論の一つとして、この不信の問題は「情報の非対称性」、すなわち、情報を持つ者と持たざる者の間の交渉や駆け引きといった力学の問題として扱われることがある。つまり、情報を持つ側はそれを自身の利益の最大化に利用しようとする動機を持つし、そのことを持たざる側も知っているがゆえに、相手に対する不信を拭えず、本来であればお互いに協力することで全体としてベネフィットは増大するはずにもかかわらず、非協力状態にとどまることによって全体としての効率性が達成されないという問題に陥る。

社会科学のオーソドックスな議論の一つとして、不信の問題は、**情報の非対称性、すなわち情報を持つ者と持たざる者のあいだの力学の**問題として扱われる。
情報を持つ側は、それを自身の利益の最大化に利用しようとする動機をもつし、持たざる側は、持つ側の動機を知っているがゆえに、相手に対する不信感を拭えず、非協力状態にとどまらざるを得ない。

図3–5　　情報の非対称性と不信

この情報の非対称性によって「逆選抜」が生じる。これは、重要な情報や欲しい情報が市場に出回らなくなることである。そうすると、せっかくいい技術あるいはシステムがあったとしても、そこに良質なデータが集まらないという問題が起こることになる。

例えば、COCOAはうまく使われれば良いシステムだったはずだが、あまりうまくいかなかった印象がある。もともとステイホームするような従順な人は登録するが、例えば集まって飲酒するような人たち、あるいは夜の仕事に従事しているようなスプレッダーになりそうな人たちの情報は最も「良質」だが、身バレするとか情報が目的外に使用される恐れが付きまとうがゆえに、すなわち、行政に対するディストラスト、不信があるがために登録しない。結局、良いシステムだったとしても、そこに集まる情報は必ずしも「良質」とは言えないものであって、したがって、評判もなかなか上がらず、みんな使用しない、という問題が起きる。これも結局のところ、情報の非対称性から来るディストラストの問題である。

したがって、まずはこの持つ側と持たざる側の情報の非対称性を克服するために、「見張りを見張る人を見張る」システムをどう構築していくかが、技術開発側に求められる。そこには、稲谷先生のおっしゃったように、法制度的なアプローチや法社会学的研究も寄与するだろう。これは、対象真正性や内容真実性を担保するための、法的、規範的なルールをどう整備していくかという問題である。あるいは、手塚先生のお話にあった、トラストアンカーをどう設計するかにも帰着するだろう。

ただ、このいわゆるデジタルトラストは、従来の政治学が仮定してきたような国家という閉じられた統治機構の中でだけ作動する問題ではないために、国家や社会の内部で伝統的に効果を持ってきた規制や強制力、あるいはサンクション（脅しも含めて）の効果は限定的にならざるを得ない。これが手塚先生のおっしゃった国際的なトラストチェーンが必要になる理由である。

そうすると、これは私の専門である国際政治学の分野に返ってくる。単一の中央政府が存在しない状況において、トラストチェーンを含めた「統治」をいかに機能させるか。この問いは国際政治学の範疇であり、グローバルデジタルガバナンスというような領域の立て方もあり得る。これは、杉村先生のおっしゃったAIの国際標準化の問題にも関わる。

単一の中央政府が存在しない状況下で、いかにトラストチェーンを含めた「統治」を機能させるか。この問いは、国際政治学の範疇であり、たとえば、**グローバル・デジタル・ガバナンス**、の視点が重要になってくる。

杉村先生のおっしゃったAI国際標準化の問題ともリンクする

図3–6　　グローバルデジタルガバナンス

今お話ししたことを踏まえ、トラストに関する基礎研究において社会科学分野の研究課題として具体的にどういうものがあり得るか、私なりに考えた3つの研究テーマ例を最後にお示しする。

- デジタル時代の「不信」の研究：　不信のメカニズムは更新される（た）のか。
- グローバル・デジタル・ガバナンスの構築（応用）：　国際標準化、グローバルな規制をいかに創設し、機能させるか。
- 身体性／関係性／中間団体がトラストに与える影響：　情報の非対称性をいかに緩和するか

図3-7　　具体的な研究課題の例（トラストに関する基礎研究）

3.3 人文・社会科学の視点から（2）

内田 由紀子

私は文化心理学を専門としている。何度か話に出た社会心理学を方法論として用いながら、一方で比較文化にこれまで長く取り組んできた。人文・社会科学の視点から今回のテーマに関連しそうなものは大きく2つある。一つは国際基準をどのようにするか。もう一つは、もう少しベーシックな社会心理学的な視点から、どういう貢献ができるかである。

国際基準に関しては、例えばAIに対する考え方の違いという話が出てきた。人が作るけれども人の手を離れてしまうかもしれないものをどうコントロールするべきかという考えに基づいてトラストを考えるような欧州的な、北米的な社会の考え方と、いったん作ったらそこにエージェンシーが発生することを受け入れる日本的な考え方がありそうで、根本的な人間観や世界観にも通じる点だろう。

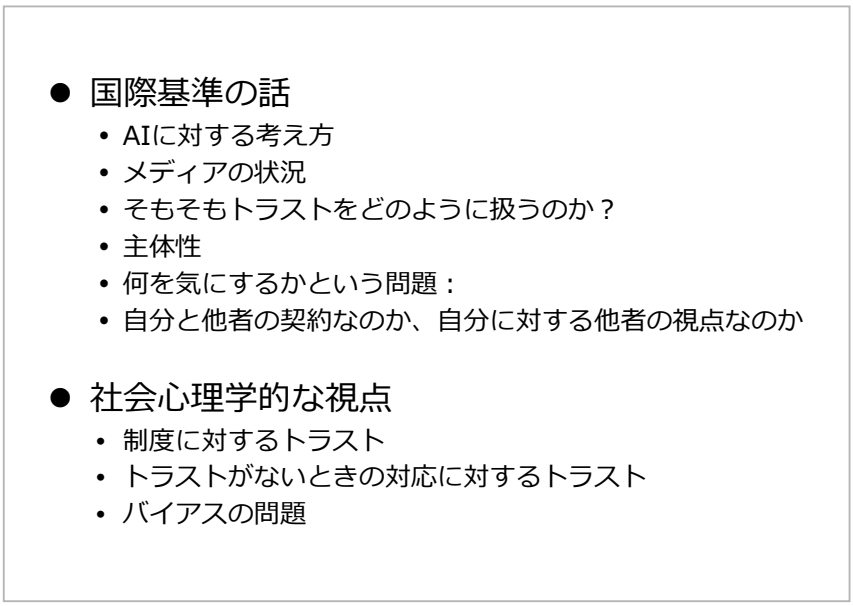


図3–8　　これまでの話からのメモ

また、メディアの状況も異なるという話があった。私は今の大統領の選挙の期間に米国にいたが、テレビのニュースも大きな新聞社も右か左かがはっきりしていて、自分と違うものを読んでいる人たちに対し「あいつらはフェイクニュースを読んでいる」といったバッシング合戦が高まっていた。

日本のメディアはニュートラルでいようとする傾向が強い。私自身が行った研究では、東日本大震災後の風評被害など、未曾有の事態が起こり記者たちもどう対応していいか分からない中で、記事の書きぶり、トーンが慎重になり、かつ、あまり自信がないということを記者たちが結構自覚していたということが調査の結果から分かった。自信がないことを開示することで信頼を担保させていこうという日本のメディアは、海外とはかなり異なると感じている。

そもそもトラストという概念をどう扱っているのかにも文化差があるだろう。また、自分が主体的な人間として正しいものに対してトラストをするという主体モデル的なトラストの在り方と、それから、制度が担保してくれていて、違反者が出たら誰かが罰を与えてくれるから、自分できちんと精査する必要がないという形でのトラストは、形態は随分異なる。そのどちらに依存するかにおいても、文化差や国の中の制度の違いなどが見

受けられるだろう。

特に日本の場合、主体性が強くないカルチャーなので、トラストできる対象かどうかを自分で決めたり、その手がかりを自分で探し当てたりすることは難しい。一方で日本では自分に対する他者の目線みたいなものがあって、周りに対してうまく説明責任を果たせる状態、周りからあまり責められない状態が担保されているときにはある程度自分発のトラストを発動できる。しかしそういうアンカーがない状態だと、主体性が発動しにくい。そうすると、決められた、所与のトラストに頼りがちになるだろう。

このように、文化心理学や社会心理学などの領域ではいろいろな形で研究が進められているので、国際基準を考えていく上で連携の手がかりは相当あると思う。

また、トラストがないときのパニッシュメントをどうするかとか、手がかりとする情報に対するバイアスの解消をどうするかなどに関しても、社会心理学とはいろいろな形で連携できるだろう。

人文・社会科学の切り口というのは横串的で、「価値」「主観」「認識」という包括的な視点を、デジタル社会のトラスト問題に関しても提供できる分野なのではないか。

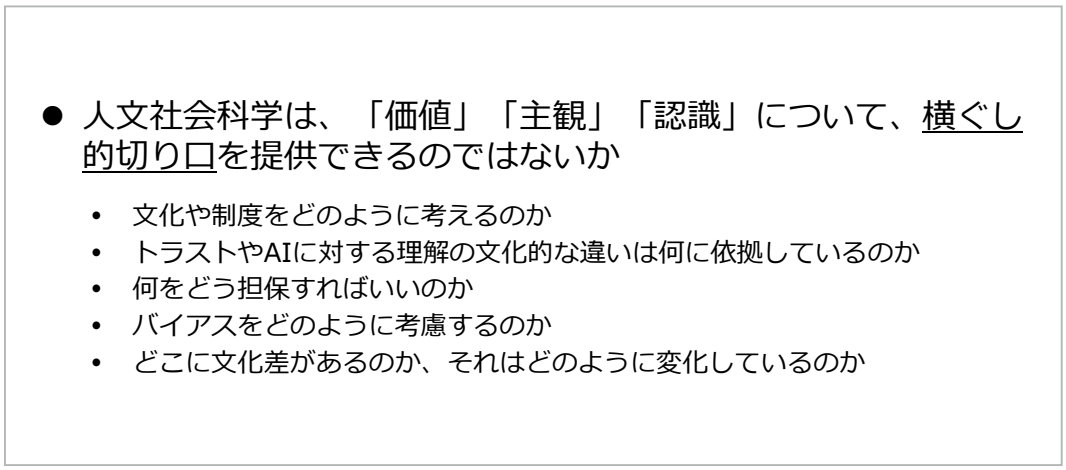


図3–9　　人文社会科学からの参画

トラストについての包括的な議論として「そもそもトラストって何だろう」「そもそも日本的な価値観って何だろう」「人がトラストするときの主体性って何だろう」といった横串的な切り口で考えるとき、人文・社会科学が提供できる知見はある。例えば、文化や制度の考え方。文化心理学の中でも、どこに文化差があってどこにないのかは重要なポイントだ。文化の変化も議論される。

文化心理学では、自然に対する視点が文化によって違うことが指摘されている。日本では自然への畏怖が強いが、欧州や北米では自然をコントロールしようとする傾向がある。それと似たことは、AIへの態度にも表れるかもしれない。日本では、AIが犯すミスや失礼を許す傾向や、AIを仲間としたいと考える傾向が他の国に比べて高いといったことも言われている。こうしたAIへの態度は、自然に対する態度やアニミズムとも奥で結びついているといった、より包括的な視点で考えることができる。

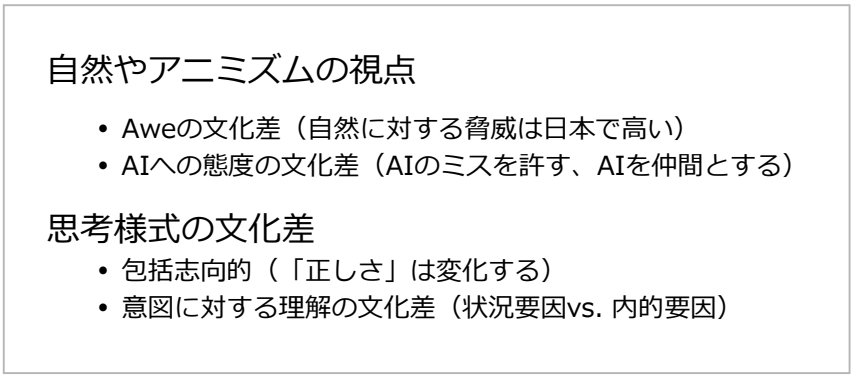


図3–10　　関連する事象として

トラストは状況判断にも影響する。この状況判断が絶対的正しさに対するトラストなのか、それとも、日本でよく見られる包括志向的な思考様式、つまり「万物は変化するものであり、今日正しいと思ったことが明日は正しくなくなるかもしれない」という価値観がある中では、正しさやトラストに対してフレキシビリティーを重視する傾向があるのではないか。また、意図に対する理解にも文化差があることが知られている。例えば誰かが悪いことをしたり、AIが悪い意思決定をしたりするときに、その内部に原因を帰属するのが欧州や北米では主流のパターンだ。しかし日本や中国などでは状況要因に帰属しがちで、たまたまこういう状況下で発生してしまったエラーだと考えることが多い。そうすると、トラストがうまくいかなかったとき、どこに原因を追及して是正するかという態度にも文化差が見られるかもしれない。

人文・社会科学の中では心理学以外の領域でも、トラストに関連する概念はさまざまな形で扱われている。私自身は最近、ウェルビーイングに関する研究にいろいろな形で参画しているが、心理学以外でも多くの人がウェルビーイングについて考えている。ウェルビーイングの絶対要件としてトラストがあるのかなどを考えると面白いだろう。

また、トラストについて重要な要素とは一体何か。それは教育、学習したり、学習機会を提供したりできるか、という視点にも広げて考えることもできるのではないか。

4 ｜ 総合討議

本章は、ワークショップ当日の総合討議パートの内容に加えて、第３章のコメントに対する質疑・議論、および、口頭での討議と並行して書き込まれたチャットでの議論、ワークショップ後に追加でいただいたコメントも含めて記載した。

産業界の視点

福島：産業界からのトラストへの期待は一致しているのか、あるいは、まだまだトラスト分野への投資には課題が大きいのか、どのように感じているか？

浦川：投資はどうしても収益を伴うところに収れんすべきという議論が強い。まだまだ一握りの企業、一握りの経営層での理解にとどまっている。

福島：6月6日・7日の世界デジタルサミット2022[1]では「デジタルトラスト ～信頼できるネット社会へ」という表題を掲げていたが、中身としては情報セキュリティーの問題が中心だった。トラストは最近バズワード的に使われている面もあるかもしれない。

浦川：私もトラストがバズワード化することは懸念している。サイバーセキュリティーに関しては理解が深まりつつあるが、ここでゼロトラストのような発想がだんだんデファクトスタンダードになっていくと、インターネット技術におけるTrusted Web化と真逆になる。

中川：ゼロトラストは重要だが、ファイアウォールで100%防ぐのはもう無理だということはネットワークの専門家の常識である。それに対して、ピンポイントで防ぐことがゼロトラストだという考え方もあるので、産業的に負荷が大きいとは言え、ある意味ではやむを得ないところもある。

ただ、その負荷が大きい場合に、もう一つの考え方としては、破られることを前提にして、すぐに復旧できるようなレジリエントな対策を提言していくことも重要だろう。それがまた経済的な負担になるが、産業界の方はどうお考えなのか？

浦川：当社では何とか防御しているが、手口は日進月歩である。ここ数年、それをパターン化し、徹底的に情報を収集、連携し、感染しかかったら直ちにたたき潰すような防御の仕組みも用いるようになってきた。インターネット技術のトラストの機能の中で、トレース機能やお墨つきを与える機能は、さらにマーケット全体で標準化が進み、企業の負担が下がるとよいと考えている。

手塚：ゼロトラストは、私が取り組んでいるトラストサービスそのものだと思う。セキュリティーの視点で見れば、ゾーン型で防御していたものがマンツーマン型に変わるということ。ゾーンが破られた後、エンティティーや人がアセット、リソースにアクセスする権限があるかの認証（Authentication）・認可（Authorization）の機能である。そのアクセスコントロールが、それぞれのリソース、アセットごとにできている環境を作ることで、究極的にはファイアウォールも要らなくなる。そのときにはポリシーコントロールの概念で、例えば社長と新入社員とハッカーでは権限が異なるということになるだろう。

そう考えると、ネットワークレイヤーで考えたゼロトラストが、ミドルウェア、さらにアプリケーションレイヤーまで広げて考えられる。ワンポリシーではなく複数ポリシーをコントロールする考え方で全てのシステムを見ていくと、トラストもマルチトラストで、それぞれにトラストの形成を見ていく必要があると考えているが、いかがか？

1　世界デジタルサミット2022
https://www.digital-summit.jp/2022/

浦川：なるほどと思える。ファイアウォールは複数の防御の一つで、それを含めた3～4層の多層型の防御をしており、情報ポリシーに従って、その多層化の分け方もしているし、ID管理の認証などの仕組みやソフトウェアに関するアップデートなどを、一網打尽というよりは、うまく最適化してやっている。まさにおっしゃるポイントと重なる。

村山：「内容真実性」はTruthだが、参加者がそれぞれ考えるBeliefも大事だろう。

浦川：政府や標準化団体がお墨つきを与えると真実になるということではない。それぞれの企業や事業体がある一定の権限を持って、お客様のデータをお預かりし、定期的にそれを更新する。データの信憑性は、利用用途や優先順位付け、データ項目によって変わるので、ここでも多層的な評価軸は必要だろう。

松本：浦川さんが「“Trust”は大前提の時代へ」と書かれているとおり、産業界的には、トラストはもう競争の話にもなっており、対応できないと企業の存続に関わる、と私は認識している。そこには、技術的な理解も必要だし、制度的な改革も行わないと日本が沈没するかもしれない。また、このトラストの考え方がデジタルトラストに移行していくと考えている。それは企業間の連携の実体が、企業システム間の連携であり、その際、信頼される側も信頼する側もシステムそのものに近く、また、今後の社会においては人の介在が最小限に抑えられたスマートコントラクトなども行われるという世界観がある。それはフェイクニュースの話などとは違うので、話がかみ合わないことがあるが、ここに向かってどう戦略を組むか、経営層も含めに認識してもらうことが必要になっている。

ゼロトラストに関しては意見が異なる。私はサイバーセキュリティーが専門だが、ゼロトラストはデジタルトラストそのものだと考えている。現在、ゼロトラストはビジネスワードとなり、ソリューション販売のうたい文句となっている側面が強いが、ゼロトラストアーキテクチャーが目指していることは、デジタルトラストにより実現する世界観と重なる。 Trusted Web自体は、リポジトリーのトラストを保つものであって、そこに接続されるスマホのようなクライアント側はゼロトラスト環境で動くことが要求されている。

小山：「トラスト」という言葉はさまざまな文脈で多義的に使われているので、「デジタルトラスト」という用語は積極的に使っていくのがよいかもしれない。そうすれば、例えばゼロトラストの話でも、純粋に技術的な領域（デジタルトラスト）と、人間が関与するそれ以外の領域を切り分けて議論することが容易になる。

ディストラスト

村山：川名先生と同様に「ディストラスト」は研究すべき分野と考える。トラスト、ディストラストは対立したものでなく、中にAbsence of Trust、トラストがない状態があるのではないか？ 以前CHI（The ACM CHI Conference on Human Factors in Computing Systems）のトラストワークショップで、eBayの参加者に「トラストの反対語はディストラストか」と聞くと「Absence of Trustだ」と答えた。ゼロトラストとAbsence of Trustが同じ概念かは不明だが、我々が求めているのは「トラストがある世界」あるいは「ディストラストがあるが安心感があればいい」とも考えられる。

川名：哲学的な問いで難しい。その領域を設定することで、これまでの政治学が解けなかった問題にチャレンジできる可能性はある。

福島：デジタル化に関係なく、もともとあるディストラストをきちんと考えるべきか？

川名：はい。例えば「分断」という問題が指摘されたが、今日の分断の原因には、デジタル技術と政治の共犯性が見られる。デジタル技術そのものはニュートラルかもしれないが、それを利用しようとする側、情報を持つ側の意図がその分断に拍車をかけている。例えばエコーチェンバーの問題は、技術的にニュートラルに生じる問題だとしても、ある種の為政者にとってその状況は歓迎的に捉えられる。トランプ時代に米国で見られた一つの現象であり、分断している場合には特定のグループに対してのみ、為政者から直に有権者に対して情報が下りてくる。それがそのグループの中で合意され、その合意された

ものを真実やファクトと呼ぶようになる。為政者にとっては非常に都合がよく、コントローラブルになってくる。政治力が考えてきた権力に対する抑制的な装置、あるいは、政治に対する不信の芽は、今日のデジタル技術、それからその先にある分断を考えるときに重要だと思う。

小山：以前のワークショップで村山先生と議論したことがあるが、トラストは、失われて初めて気づくもので、ありありと実感するという方が例外的だ。ディストラストからトラストを考えるという方向性の方が適切かもしれない。

トラストの３側面の統合・複合の方向性

福島：今回の提言では、技術的な観点からトラストを３つの側面で整理して、それを統合・複合することでトラストがきちんと作られた世界に持っていく方向を考えている。そのような３つの側面を統合していくところに関して、あるいは、そのための情報技術的な観点でのこれまでの取り組みと人文・社会科学系の考え方をうまく融合して進めていくことに関する議論はないか？

松本：対象真正性を手塚先生、内容真実性を藤代先生、振る舞い予想・対応可能性を杉村さんがお話されたが、今日の話としては、３つの分け方には違和感を持った。杉村さんは、トラスト自体ではなく、Trustworthinessの話、AIの信頼性の話をしておられたと思う。

Trustorがどうやって Trusteeを信頼できるかというトラストのメカニズムの話はされていない。逆に、手塚先生は、信頼する側であるTrustorが検証可能になることが重要だと言っており、これは、トラストのメカニズムとその基盤の話になる。これは、対象真正性と振る舞い予想・対応可能性の違いの話ではない。ここに違和感があった。

手塚先生の話は、トラストサービスは対象真正性とされたが、対象真正性に限らず、デジタル的に、リアルタイムに検証できること、これはトラストのメカニズムの話ではないか。内容真実性も振る舞い予想・対応可能性も、何らかのデジタル証明ができれば、デジタルに検証できる。例えば欧州のeIDAS 2.0では、自然人、法人の属性証明をやろうとしており、対象真正性から内容真実性にどんどんシフトしている。一番元となっているのは対象物としてヒトそのものだが、その属性の方に行っている。なので、Trustworthinessとトラストの話は切り分けなければいけないと思う。 Trustworthinessは、Trusteeの信頼性を言っている。その信頼性が、Trustorが検証可能な仕組み、これは、トラストのメカニズムであり、これを作ろうとしているのがトラストサービスの話。トラストに関する分類は、この３つの軸だけでは説明できていないと思った。

福島：トラストの３側面を挙げたのは、それらをばらばらに扱うのではなく、統一的に扱えるような枠組みを考えていくべきと言いたかったためだ。手塚先生、藤代先生、杉村さんの取り組みは、現状、対象真正性、内容真実性、振る舞い予想・対応可能性にそれぞれの重点があるが、それぞれがそれしかやろうとしていないというわけではないと承知している。手塚先生の質疑でも、枠組みをうまく拡張して、カバーがされるところを広げていく方向を示唆されたと思う。松本さんの言うような検証可能なトラストのメカニズムが３側面を統一的に扱えるように発展するというのが、一つの方向性だと思っている。

一方、トラストの３つの側面を複合的に考えるのは、法やガバナンスの立場が近いように思える。そのような取り組みは、まだ進んでいないのか、いろいろな取り組みが既にあるのか、世界的な取り組み状況や取り組み方に関して意見をいただきたい。

大屋：法制度的な観点では、総論は粗い話しかできない。ハードローのようにリジッドなガバナンスからアジャイルへと、今日、稲谷さんがされた話は総論としてはやっていて、できる。しかし、アジャイルガバナンスで巨大エンジニアリングがやれるわけではない。例えば、原発はハードローで、ウォーターフォールモデルで管理しないといけない。個別具体的な問題領域を設定し、当事者の知識レベルや、そこでAIがどの程度働けるか、などの問題がフィックスされれば、ガバナンスとしてどういうことが必要であるかの議論はそれなりに細密にできる。

法制度自体は、近代的な自律的個人モデルをかなりの程度諦めている。消費者法があり、行政法がいろいろやっている中で、個別に具体的な分野でさまざまな道具を作ってガバナンスする。総合的にそれらを全部知っている人がいるわけではない。技術的側面とのすり合わせが行われているかは分からないところもある。

稲谷：大屋先生のおっしゃるとおり。各パーツを全部統合してこういう方向に持っていくとの議論はない。各論はそれなりにやっている。しかし、それぞれの領域や階層を横断・縦断する研究はやはりできていない。今回の対象真正性、内容真実性、振る舞い予想・対応可能性も、それぞれの法分野で違うものとして議論されてきた論点が含まれている。

例えば、刑事法で見ていくと、対象真正性は、偽造の問題として典型的に議論されてきた問題。有形偽造で、明らかに処罰性があるだろうと議論されてきた。内容真実性は、特定のケースに関して名誉毀損の問題を考えなければいけないが、そうでなければ、基本的には別にいいと議論されてきた。むしろ憲法との関係で議論される傾向がある。振る舞い予想・対応可能性は過失犯の問題で、それはそれで専門的にやっている。

全体として、それぞれの規範を組み合わせて何かを実現したいとの視点は多分ないが、それぞれに細かい議論はしている。ただ、技術の人と対応しながらやっているわけではない。研究プログラムを考えるなら、第０層から第４層まで、例えば個別的な１個の論点でもいいが、何か思想的なところ、背景的なところから技術を通して、具体的なアプリケーションまで一貫して考えるプログラムを増やすことが重要なポイントだと考える。

浦川：問題領域が４次元ぐらいに複層化していると感じる。４つの次元について、例えば、これはTrusted Web 推進協議会で検討している、これは経産省のデジタルガバナンス検討会でやっているとなっているのが、今後の推進体制の議論が発端だったと思う。ガバナンスはこういったことを推進する上で、規制的な発想ではなく、ポジティブに後押しをするアクセラレーターとしての役回りをデジタル庁ほか行政の方々に担っていただきたい。アカデミアや産業界の心ある人のために、こういうJSTのような場が膨らんでいき、官庁がばらばらにやることを内閣官房に束ねてもらう。このように全く違う専門の方々が一つの抽象的なトラストというキーワードで議論をすることが本当に貴重だ。

トラストと検証についての考え方

稲谷：私もトラストという概念の位置付けが、分からなくなってくるところがあった。特に横串を通していくとき、これを使って何をしたいのかが重要。もともとトラストは、とりわけ取引におけるトラストは、未知の人とうまく付き合う戦略を確保する文脈の中で存在している。内田先生のお話は、「知っているものにする」というアプローチかと思う。未知のものを無理に信頼してもらう方向性を選ばず、とりあえずAIによって知っているものにしていくのなら、自分で情報費用をかけて検証する必要はない。検証してもらって安全性が確保されているなら、使ってもらえる状態が生まれる。どこをゴールにするのかが明確になると、横串の通し方をどう設計して、それと合わせて技術をどういう形でアピールしていくかも変わってくる。どちらの世界観がいいかを今決める必要はないが、２つの方向性のどちらに行くかはポイントである。どちらに行ったとしても、川名先生が指摘された政治権力の濫用のリスクをどう押さえ込むかは問題になる。ただ、大きな分け方として２つの方向性があり、特に人文・社会と組んでやる話になると、そのどちら側をどういう形で組み合わせてどう見せていくかが、議論を深めていく上で重要なポイントになる。

手塚：直交関係ではないと思っている。基本的にデジタルを対象に考えたい。デジタル空間とフィジカル空間は違う。（フィジカル空間は）一度相手が分かったから、その後はFace to Faceでやってトラストのレベルが上がり、チェックする項目を減らしていく。サイバー空間にそのメカニズムを入れるとすると、パブリックトラストとプライベートトラストの概念が出てくる。一般のパブリックトラストはゼロからやっ

ていく。この人は非常に高いレベルのトラストだからここを省略するなどの、属性情報を入れたメカニズムを入れるのは難しい。サイバー空間では自動化によってナノ秒でやれるので、基本的に性悪説型で全部をシステムに固める考え方だと思う。その後、必要に応じてメカニズムは付け加えていけばよい。基本は全部ゼロ。パブリックなサイバー空間、ワールドワイドにみんな入ってくる空間では、その考え方でまずはやっていく。その後、メカニズムを入れていくのはそれぞれ工夫が要る。直交関係では特にないと思う。

稲谷：今のポイントは非常に面白い。性悪説で全部やっていくのは、典型的に信頼がない社会のメカニズム。逆に、今まで時間をかけて大量に調べないといけなかったものがナノ秒で処理できると、取るべき戦略は今まで信頼を作っていこうとした戦略と違うかもしれない。

サイバー空間はつながっている。一つのシステムやサービスを提供するとき、複数の人がつながりながら展開をして、そこの関係性で問題が生じることがある。そうすると、見知らぬ人同士の一回的な取引を促進するために、信頼を醸成するシステムとは違うと言える。むしろ、信頼できない人同士がうまくやっていくためのシステム。今まで、知らない人同士で信頼できている世界とできていない世界があって、知らない人同士が信頼できる世界にどう近づけるかとの話がされてきた。そうではないところが面白いと思う。

手塚：もう一歩進めると、ハッカーとの関係が出てくる。だから「信頼できているやつも信頼できない」との行動に変わる人がいる。スノーデンのような例を食い止めるメカニズムも本来はサイバー空間の中には入れていくべき。やはりまずゼロの状態で常にやる。そのメカニズムを一度作れば自動化によってナノ秒でコンピューターが処理する。一般社会の価値観でものを考えず、サイバー空間上でどうあるべきかの価値観で考える時代になってきている。

村山：ベリファイされても何となく不安という状態も人間ならばある。デジタルの世界は、ベリファイできたならトラスト度が上がるのか、自動的になるのか？　AIは人間が何となく不安なところをフォローしてくれる思う。

福島：私の発表の際に、多面的・複合的にベリファイする必要性を述べた。各情報からこれぐらい大丈夫という程度付きの判断材料がそれぞれ得られて、それらを組み合わせて総合的に判断するというのが方向性だと思う。たくさんの判断材料をもとに多面的・複合的にベリファイするのは人間には大変なので、AIよる判断の自動化や人間が判断しやすいように整理して見せることなどが必要になると思っていて、これは今の話に近いように思う。

大屋：検証の対象が広がっている背景として、人為的にコントロール／設計可能なシステムの範囲が拡大しているという問題がある。物理的に自分の顔形を変えるのは極めて難しいので、顔写真が個人識別の手段として多用されてきたわけだが、アバターを使うコミュニケーション手段では、そこが人為的に操作可能になり、対応策を我々が設計すべき理由が生じる。

見張りの見張り問題

藤代：フェイクニュースの世界では、ファクトチェックをする人や組織をどうファクトチェックするのかという議論が既にある。見張りの見張り問題は、ファクトチェック団体のような善意のプレイヤーが、ガバナンスを受け入れるメンタリティーに乏しいという点も関わってくる。

大屋：善かれ悪しかれ、我々は合理的な意思決定に基づいて、あるシステムを利用するかどうか決めるのではなく、なんとなくとか、昔から使っているからといった理由でシステム利用の可否を決めてしまっている。こんにゃくゼリーが（喉に詰まりやすいことが）問題になったのに餅が許されているのはなぜかというように。である以上、社会的に必要なシステムであれば、情報発信による人々の合理的な意思決定に期待するのではなく、圧倒的なメリットで目をくらませて、リスクのことを忘れさせるというソリューションも考えられる。

中川：圧倒的なメリットでの目くらましは、まさにGAFAの戦略だと言える。それが若干でも正しく機能するというなら、利用者の評判を落とすというポイントが働くかどうか。その場合は、代替システムがないと効果的でないので、競争法や独占禁止法が重要になる。そういう意味ではEUのGDPRはある種の成功例かもしれない。

藤代：政治権力とデジタル技術の共犯は、メディアリテラシーのパラドックスの鍵である。「私は正しく読み解いている」という反応、もしくは「読み解けないのは愚かだ」というマウンティングと自己責任論が進むだけになる。

ガバナンスの評価

佐古：ガバナンスの評価のための評価関数はどうなっているのか。社会にとって良いかどうかの評価というのは「Aさんにとっては良いけれど、Bさんにとっては？」となりそうに思う。

大屋：これは主に政府や社会の視点なので、幸福であれGDPなどの指標であれ、集計量にならざるを得ないと思う。もちろん、単に集計最大化だけでなく特定個体の値について、例えば人権保障など、一定の閾値を割り込まないようにするというサブルールは必要である。

佐古：どのようなデータから集計値をとるかという点の透明性や、その妥当性の議論も重要だと思う。その計算をAIにさせる将来になるかもしれない。

稲谷：評価関数に関しては、各主体の目標設定との関係で行われるイメージを持っている。例えば、自動運転サービスを提供する企業ならば、事故率などは指標になるかもしれない。その際、そのサービスによって影響を受ける人々と協力しながら、指標を決定する仕組みを導入するべきだというのが、マルチステークホルダーアプローチの趣旨になる。これは企業にとっても、事業をサステナブルにする観点から重要だと思う。したがって、この地域に住んでいる人の中で、納得いかないということであれば、設定された指標について、自らの資格で異議を申し立てることができるし、そこで民主主義的な議論が生じることになる。その上で、問題となる指標がより根本的な価値に関係するようなもので、ハードローの妥当性が問題となるようなケースであれば、人権などを含むさまざまな基本的価値との関係で議論を進めていくことになるし、司法のようなレガシーシステムも重要な意味を持つと思う。

佐古：マルチステークホルダーアプローチで評価関数を決めた後は、それに基づいてアジャイルにガバナンスをアップデートする方向性だと理解した。ただ、ガバナンスがアップデートされるたびに、評価関数を見直していたら、なかなかアジャイルにならないという課題はありそうである。

トラスト形成における中間団体

川名：私は結局のところ、人々の不信を前提に世界を見ている。しかし、トラスト形成における中間団体の役割に期待を見いだせるかもしれない。つまり、社会においてフェイクを濾し取る、濾紙のようなものは何かを考えるとき、例えば政治学分野であれば、ポピュリズム研究のなかでも、とくに「中間団体」の研究とのリンクがあり得る。本日冒頭のCRDSの発表にも、権力による情報の濫用、恣意的な運用を回避するために分散型のシステム、という話があった。それに関連すると思うが、今日のデジタル社会がもたらすトラストの問題は、中間団体、すなわちコモンセンスを前提にした、帰属集団が失われていることにあると考えられている。会社や労働組合、教会や町内会、PTAしかり。それらが、権力と市民の間にあって、コミュニケーションのクッション、フィルターの役割を担っていた。しかし、今はそれが失われ、またSNSなどを通じて、権力あるいは行政、インフルエンサーと個人が、直接コミュニケーションするようになっている。かつての社会にあった緩衝材はない。そのことは、権力、あるいは不正確な情報を発する側にとっては非常に都合が良い。したがって、デジタル社会におけるトラストの構築は、テクノロジー側の解決策と同時に、社会の側の、コミュニティー側の体制も、周到に準備しなければならない、ということである。中間団体がなければ、結局は、情報と個人が、直接接続される、

という構造自体は保存されてしまうことになる。社会の側の緩衝材、中間団体、コミュニティーに関する研究も必要だろう。

藤代：中間団体はミドルメディア論にも接合できそうだ。今は技術を味方につけたものが、人を集めて別の中間団体を攻撃するので、中間団体がどう涵養されるかの設計は重要そうに思う。

大屋：現代における一つの手段はプラットフォーマーだが、そこに競争が乏しいということを含めてTrustworthinessがあるかが論点になる。中間団体は国家より小さいか、もはや、かつてのカトリック教会と国家の関係のようなものになっているのではないかという問題提起は必要かと思う。

稲谷：私もプラットフォーマーの問題は大きいと思う。彼らに対する正負のサンクションの設計を市場メカニズムとどう組み合わせるのかが、鍵になると思う。

中川：中間的ないし中立的な媒体として新聞がどのような将来像を持つべきかを議論したことがある。紙の新聞が消滅したとき、戦場など危険な場所に行ってファクトを発信してくれるジャーナリストをサポートする、ないしは種々の意見やファクトを総合的に見て人々に伝える編集的な役割が必要になる。しかし、これが果たして経済的に成立するかどうかが心もとない。

大屋：現状ではプラットフォーマーが伝統的メディアのニュースにただ乗りできていることが問題で、支払いが必要になれば、Netflixのように自分でファンドすることを考えるかもしれないし、オーストラリアのニュース税のような話もある。

トラストの状況変化と保険

福島：推進策について2つの観点でご意見いただきたい。一つは、分野間の連携、特に情報系と人文・社会系も含めた分野間の連携の形でトラスト研究を推進する上のやり方について。もう一つは、国際的にリードすることも見据えた日本のこれまでのポジションや今後の取り組みについて。

中川：トラストをどう確立するかという話がこれまでされた。トラストは一度確立すればずっとトラストし続けられるというようなイメージが多い。今後の話として、今までトラストしていた状況や相手に対する条件が変化して、それにどう対応するかがある。誰かがなりすましたり乗っ取ったりしたのでその人がトラストできないようになったのか、本当にトラストできない考えに変えたのか。技術的に難しいし、人文・社会科学的な点から見ても難しい。また、ディストラストな状況になったときに発生する損害をどう補償していくかの仕組みも議論されていない。

松本：中川先生のお話された「トラストしていた状況の変化」だが、自動車や医療デバイスに関して、現在、法律上は、型式証明、例えば車種に証明を与え、出荷検査をして道路を走るといった規制の枠組みであり、これが、これらのトラストを確保する仕組みでもある。しかし、自動運転車、AI医療デバイスなどを念頭に置いた場合、サイバー攻撃の対応や機能追加もオンラインでのソフトウェア更新といった形で行われることが想定され、そこでは、「トラストしていた状況の変化」も生じる。サイバー攻撃の対応のソフトウェア更新が行わなければ当然トラストも低下していく。そのため出荷後も意図どおり動いていることを証明しなければならない方向に向かっており、トラストの対象となる自動車、医療デバイスの稼働期間、サービス期間中に責任を持つという法制度が作られつつある。例えば医療デバイスの認証を行っている米国FDA（Food and Drug Administration）もそういうことを大分以前から取り組んでいる認識している。

稲谷：責任の問題が出てきている。いろいろな形で今までと責任の境界線が変わってきている。中川先生も指摘したいくつかの主体が関係する場合や、松本さんが指摘した製品の完成や流通の考え方が変わってくる問題がある。さらに、リアルでアップデートされて、製品自体も変わっていく状態になっている。責任を上手に分配する方法を考えていかないと、必要なアップデートをやらなくなり製品の信頼が下がっていく。全体として良いAIシステムを作っていくために、継続的に作り続けていくためにどうするのかは視点として入れてよい。人々に寄り添う形のシステムを作っていくための視点である。日本的な

寄り添って一緒に変わっていくモデルに近い。その議論は正面からされてない。問題提起はされて着手はしているが、掘り下げてやっているわけではない。国際的にルールが決まっているわけでもない。技術側で何ができて、法制度側で何ができて、文化的にどういうものがアクセプタブルなのかを共同研究するのは、面白い方向だと思う。

中川：稲谷先生のご指摘とおり、トラストが崩壊しないために制度的にどう担保するかが重要。それに対して保険がある。ある人が期待を裏切られて損害が発生しても、保険で多くのことはカバーでき、かつ、その人は次から保険に入れなくなるなど、うまくできた仕掛けである。それをトラストから生じる責任や損害に対してどう適用するかはこれから。現実問題として、例えば自動運転では考えざるを得ない問題として出てくる。重要、かつ、近い未来の問題である。

稲谷：保険の問題は中川先生がお話されたとおり重要だ。保険会社に期待される役割も、保険会社自体の業態も変わってくると言われている。保険会社がデータを取ってモニタリングをすると、今までとは違うレベルでモラルハザードを監視することができる。保険会社が入ることによって、全体最適に近づきやすい状況を生むことができるとも言われ、産業的な要請も大きいと思う。現実的に、既に自動運転と保険に関して本も出ているが、そこで言われていた話から条件も変わっているので、議論してみる余地はあると思う。

方策のイメージ

参加者：今後、必要な方策に関して、研究拠点とファンディングが挙げられていたと思う。拠点であればどのようなイメージのものが好ましいのか？　例えば、理研のAIPセンターのELSI研究グループのようなものをベースに考えているのか、あるいは、海外の先行例となるような研究拠点をイメージしているのか？　また、研究費は、例えば、さきがけのような基礎研究を行うファンディングがよいのか、それとも、出口を意識して具体的な社会制度の案を作るようなバックキャスト型がよいのか、あるいは、これらを混ぜたようなファンディングがよいのか、設計のイメージは何かあるか？

福島：本日冒頭で図1–9のような素案を示したが、これをたたき台として、皆からご意見をいただきたい。例えば、トラストの共通基盤研究の拠点は、ELSI/RRIの研究拠点と位置付けが近いかもしれない。ファンディングプログラムに関しては、今回は大きな絵を描いているところで、そこから複数のファンディングプログラムが立ち上がっていく形が考えられる。具体的トラスト問題解決のための目標を定めて取り組むという切り口も考えられれば、基礎的な基盤の研究開発という切り口も考えられる。まずは、連携促進の場作りから始めて、その中で具体的な目標が設定されて、そこにファンドを付けるような流れが考えられるが、ご意見をいただけるとありがたい。

デジタル社会ならではの文化差および匿名性

住田：内田先生の話にあった文化差について、デジタル社会ならではの文化差の研究は進んでいるか？　日本ではネット上で匿名にする人が多いようだが、それによって主体性が発揮しやすいなどもあり得るように思った。

内田：ソーシャルメディアに関連する比較の心理学研究は少しずつ出てきている。感情の共有の仕方やネットワーキングの仕方の違いといったものが現在は主流になっている。匿名になったときに出てくる感情は主体的になっているようにも見えつつ、逆に普段主体的に抑制している感情を全く制御せずに出しているともとらえられるので、一概には言いづらい。

大屋：フランス革命期に、選挙における投票は公開での発声と書面による秘密投票のどちらが正しいかという論争があった。自律的な個人であれば自らの選択に責任を負うべきだという立場からのスジ論としては、発声投票に勝ち目がありつつ、結果的に秘密投票が現代に至る原則となったのは、人間には公開の場で言えない本音があるという見方によるものだと考えることができる。トレーサビリティーのない発言

手段には、一方で誹謗中傷や感情むき出しの発言を許容してしまうという問題がありつつ、内部告発などの機会を確保するために必要だという見通しもある。両者のバランスをどう取るかということが、言論の自由とか社会全体のガバナンスの観点からは必要になってくる。

佐古：トレーサビリティーがあっても、誹謗中傷や感情むき出しの発言が減らなかった、という韓国の社会科学の実験があり、その後、いろいろ追試されて確認されていると聞いている。

小山：ロボット・AIの許容度合い関する日英の比較研究で、若い世代では差が出なかったという研究もある。世代差の研究も重要になってくるかもしれない。

人文・社会科学の参画・連携の進め方

内田：いろいろな形でプラットフォームを形成して、協働的な枠組みを支援していただく場所は必要。今日のようなディスカッションができるだけでも非常に大きい。一回で終わらせずに継続的にやっていくためにJSTなりがサポートする。プラットフォーム作りが重要だと思うと同時に、人文・社会科学系の中でも、科学技術や理系の先生方と組むことで自分たちの研究が発展するメリットを感じられることが大切。ELSIの取り組みに関して、科学技術の側で出来上がったプログラムを人文・社会科学がフォローアップするような役割を担う形態が先行している。一方で、こうしたコンセプトを立ち上げるというところから人文・社会科学の人が一緒に入って、そこの定義はどうなんだとか、人間ってこれに関してどんなふうに考えるのかということを、最初のボトムアップベースのところから一緒に作り上げていく形での参画が、お互いにとってWin-Winだと思う。こういう形の議論の取り組みというのを続けていっていただきたい。

小山：昨年のセミナーシリーズのような分野横断の議論の場を設けて、そこから共同研究のテーマを育てていき、その段階でグラントを取りに行くというのは一つの方法だとは思う。一方で、人文・社会科学の研究者は、情報科学の研究者と比べて数自体が少ないし、他分野どころか自分野内ですら共同研究の経験も少ない人が多い。トラストの問題は差し迫ってきているものがあるので、早く立ち上げる必要があるという意味では、情報科学の既に動きつつある研究テーマを主体にグラントを獲得して、それに人文・社会科学の研究者が協力する形が先に動きやすい。いろいろな手段を並行して進めていくのがよいのかもしれない。

付録　ワークショップ開催概要

開催日程：2022年6月11日（土）13:00～17:00（4時間）
開催形態：オンライン（Zoomミーティングを使用）

アジェンダ：

13:00–13:30［30分］開催挨拶・開催趣旨・提言骨子
　木村 康則・福島俊一（JST CRDS）
13:30–15:10［100分］技術開発・制度設計の状況・課題・方向性［100分］
　手塚 悟　：対象真正性の視点
　藤代 裕之：内容真実性の視点
　杉村 領一：振る舞い予想・対応可能性の視点
　稲谷 龍彦：法制度・ガバナンスの視点
15:10–15:20［10分］休憩
15:20–16:00［40分］社会・産業および人文・社会科学の視点
　浦川 伸一：産業界の視点
　川名 晋史：人文・社会科学の視点 (1)
　内田 由紀子：人文・社会科学の視点 (2)
16:00–16:55［55分］総合討議
16:55–17:00［5分］まとめ・閉会挨拶　木村・福島（JST CRDS）

招聘参加者：

	氏名	所属・役職
発表者	手塚 悟	慶應義塾大学環境情報学部 教授
	藤代 裕之	法政大学社会学部 教授
	杉村 領一	産業技術総合研究所 情報・人間工学領域 上席イノベーションコーディネータ
	稲谷 龍彦	京都大学大学院法学研究科 教授
コメンテーター	浦川 伸一	損害保険ジャパン株式会社 取締役専務執行役員
	川名 晋史	東京工業大学リベラルアーツ研究教育院 准教授
	内田 由紀子	京都大学 人と社会の未来研究院 教授
議論参加者	犬飼 佳吾	明治学院大学経済学部 准教授
	大屋 雄裕	慶應義塾大学法学部 教授
	小山 虎	山口大学 時間学研究所 准教授
	佐古 和恵	早稲田大学基幹理工学部 教授
	中川 裕志	理化学研究所 革新知能統合研究センター 社会におけるAI利活用と法制度チーム チームリーダー
	松本 泰	セコム株式会社 IS研究所 ディビジョンマネージャー
	村山 優子	津田塾大学 数学・計算機科学研究所 特任研究員

聴講登録者（招聘者・チームメンバー以外）：
　文部科学省　1名　　経済産業省　1名　　NEDO　1名　　JST関係者　12名　（計15名）

ワークショップ企画・報告書編纂メンバー

総括責任者	木村 康則	上席フェロー	CRDS システム・情報科学技術ユニット
リーダー	福島 俊一	フェロー	CRDS システム・情報科学技術ユニット
メンバー	加納 寛之	フェロー	CRDS 科学技術イノベーション政策ユニット
	上村 健	フェロー	社会技術研究開発センター 企画運営室
	住田 朋久	フェロー	CRDS 企画運営室
	高島 洋典	フェロー	CRDS システム・情報科学技術ユニット
	戸田 智美	フェロー	CRDS ライフサイエンス・臨床医学ユニット
	花田 文子	フェロー	CRDS 企画運営室
	福井 章人	フェロー	CRDS システム・情報科学技術ユニット
	的場 正憲	フェロー	CRDS システム・情報科学技術ユニット
	茂木 強	フェロー	CRDS システム・情報科学技術ユニット
	山本 里枝子	フェロー	CRDS 企画運営室
	若山 正人	上席フェロー	CRDS システム・情報科学技術ユニット

科学技術未来戦略ワークショップ報告書 **CRDS-FY2022-WR-05**

トラスト研究戦略

～デジタル社会における新たなトラスト形成～

令和 4 年 9 月　September 2022
ISBN 978-4-88890-813-9

国立研究開発法人科学技術振興機構　研究開発戦略センター
Center for Research and Development Strategy, Japan Science and Technology Agency

〒102-0076 東京都千代田区五番町 7 K's 五番町
電話　03-5214-7481
E-mail crds@jst.go.jp
https://www.jst.go.jp/crds/

ISBN 978-4-88890-813-9

FOR THE FUTURE OF
SCIENCE AND
SOCIETY

CRDS

https://www.jst.go.jp/crds/