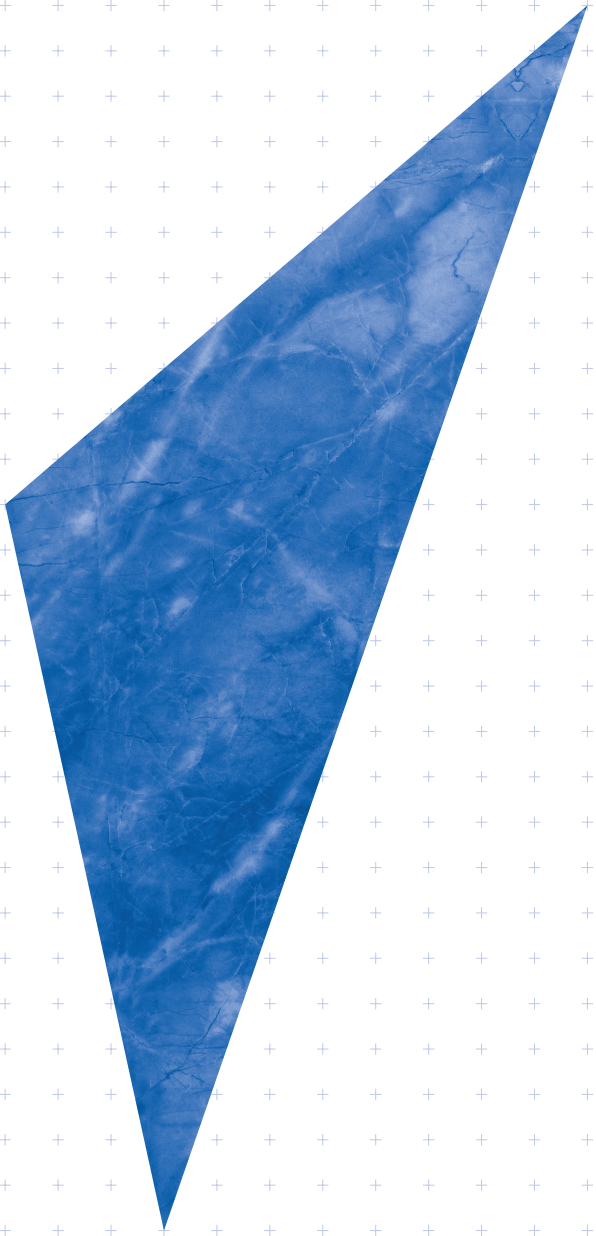
CRDS-FY2022-SY-02

公開シンポジウム報告書

デジタル社会における新たなトラスト形成

～総合知による取り組みへ～

2023 年 1 月 10 日（火）開催



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、2023年1月10日に開催した公開シンポジウム「デジタル社会における新たなトラスト形成～総合知による取り組みへ～」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である。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。トラストは協力や取引のコストを減らしてくれる効果があり、人々の活動を拡大し、ビジネスを発展させ、ビジネスの生き死にを左右する要因にもなる。

しかし、デジタル化の進展につれて、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への依存が高まり、だます技術も高度化した。その結果、デジタル社会と言われる今日において、顔が見える人間関係や人々の間のルールに支えられた「旧来のトラスト」だけではカバーされないケースが拡大し、社会におけるトラストの働きがほころんできている。この問題は、自動運転車、AIエージェント、コミュニケーションロボット、メタバースなどの新技術・新サービスの社会受容を左右し、フェイク・偽装・なりすましなどによる詐欺・犯罪の懸念を高める。

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」をテーマとした調査・検討を進め、2022年9月に戦略プロポーザルを公開した。このプロポーザルでは、上述の問題に対処し、不信・警戒を過度に持つことなく幅広い協力・取引・人間関係を作ることができ、デジタル化によるさまざまな可能性・恩恵がより広がるような社会を目指して、「デジタル社会における新たなトラスト形成」の研究開発推進を提言した。

今回の公開シンポジウムは、その研究開発推進に向けて、戦略プロポーザルの内容をたたき台として、①デジタル社会のトラストに関する問題意識を広く共有したい、②トラストを軸とした総合的な研究開発ビジョンの下絵を提示して議論を深めたい、③総合知による取り組みの必要性を共有して分野横断で継続的に議論できる場作りにつなげたい、と考えて企画したものである。そのため、まず第1部では、戦略プロポーザルの概要を紹介した。

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。第2部では、これらトラストの3側面のそれぞれに対して研究開発の状況・事例を紹介していただいた。

今後目指すべき方向性として、トラストの3側面を多面的・複合的に扱っていくべきであり、そのために、分野横断の総合知による取り組みが必要である。そこで第3部では、情報系、人文・社会系、具体的な応用分野など、さまざまな分野でのトラスト研究の動向や事例を紹介していただいた。その上で第4部の総合討議では、さまざまな分野・側面を横断するような議論を通して、総合知による取り組みや、その下地となる分野横断的な議論・知見共有の必要性が示された。

今回の公開シンポジウムはキックオフ的な位置付けであり、今後も分野横断の議論が持てるような場を設けていきたいと考えている。

目次

1 ｜ 戦略プロポーザルの紹介

福島 俊一

シンポジウム開催の背景・意図

JST CRDSの俯瞰・戦略提言活動の一環で、2021年6月から2022年9月にかけて「デジタル社会における新たなトラスト形成」をテーマとした調査・検討を行った。その間、15回のセミナーと2回のワークショップを開催し、50名超の有識者インタビューを通して議論し、知見を集積して、戦略プロポーザルをまとめた[1]。

これまでのセミナーやワークショップではクローズド形式の議論を行ってきたが、今回は公開シンポジウムとして開催する。その意図・狙いは、研究開発の推進に向けて、戦略プロポーザルの内容をたたき台として、①デジタル社会のトラストに関する問題意識を広く共有すること、②トラストを軸とした総合的な研究開発ビジョンの下絵を提示して議論を深めること、③総合知による取り組みの必要性を共有して分野横断で継続的に議論できる場作りにつなげることである。

トラストの役割と問題意識

トラスト（信頼）についてさまざまな定義があるが、おおむね相手が期待を裏切らないと思える状態と考えてよいだろう。リスクはあるのだが、トラストすると、安心して迅速に行動・意思決定ができる。取引や協力のコスト減らしてくれて、ビジネスにとっても非常に重要なものだ（図1–1）。

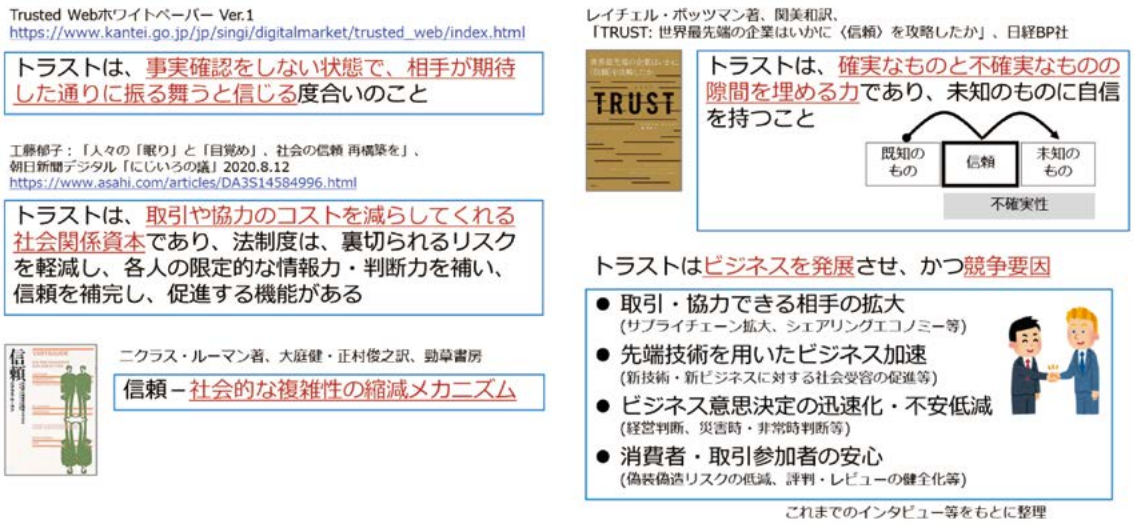


図1–1　　トラスト（信頼）の役割

1　俯瞰セミナー＆ワークショップ報告書（2021年7-10月開催、2022年2月発行）「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～」https://www.jst.go.jp/crds/report/CRDS-FY2021-WR-05.html
科学技術未来戦略ワークショップ報告書（2022年6月開催、9月発行）「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～」https://www.jst.go.jp/crds/report/CRDS-FY2022-WR-05.html
戦略プロポーザル（2022年9月発行）「デジタル社会における新たなトラスト形成」
https://www.jst.go.jp/crds/report/CRDS-FY2022-SP-03.html

しかし、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ったこと、複雑な技術を用いたシステムに依存するようになってきたこと、だます技術も高度化してしまったことなどが原因と考えられる。そこで、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を作ろうというのが、戦略プロポーザルが目指すところだ（図1–2）。

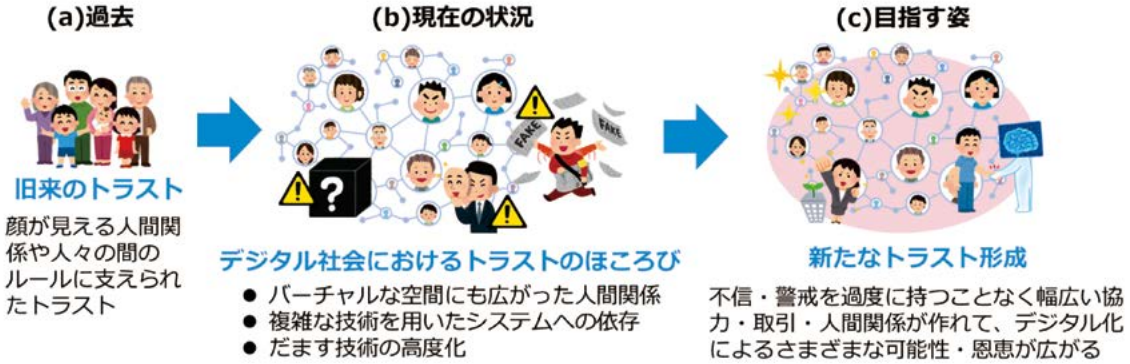


図1–2　　デジタル社会におけるトラストのほころびと目指す社会の姿

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1–3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）技術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、新たな様相を示しつつある。

トラスト研究の状況と目指す姿

トラスト研究はこれまで、人文・社会科学分野の基礎研究から、ビジネス・社会実装と連動した情報科学分野の技術開発、医療・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。しかし、それらの間で知見共有・連携はほとんど見られず、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1–4に示したようなBeforeからAfterの形に変えたい。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したい。

現在のトラスト研究がばらばらに進んできたように見えるのは、トラストに対する異なる側面・切り口から研究開発が進められてきたためと思える（図1–5）。これまで行ってきた議論から、トラストは三つの側面に分けて捉えることができると考えている。一つ目は「本人・本物であるか？」という対象真正性、二つ目は「内容が事実・真実であるか？」という内容真実性、三つ目は「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・対応可能性である。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になる。

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1–3　　デジタル化の進展が生んだトラスト問題

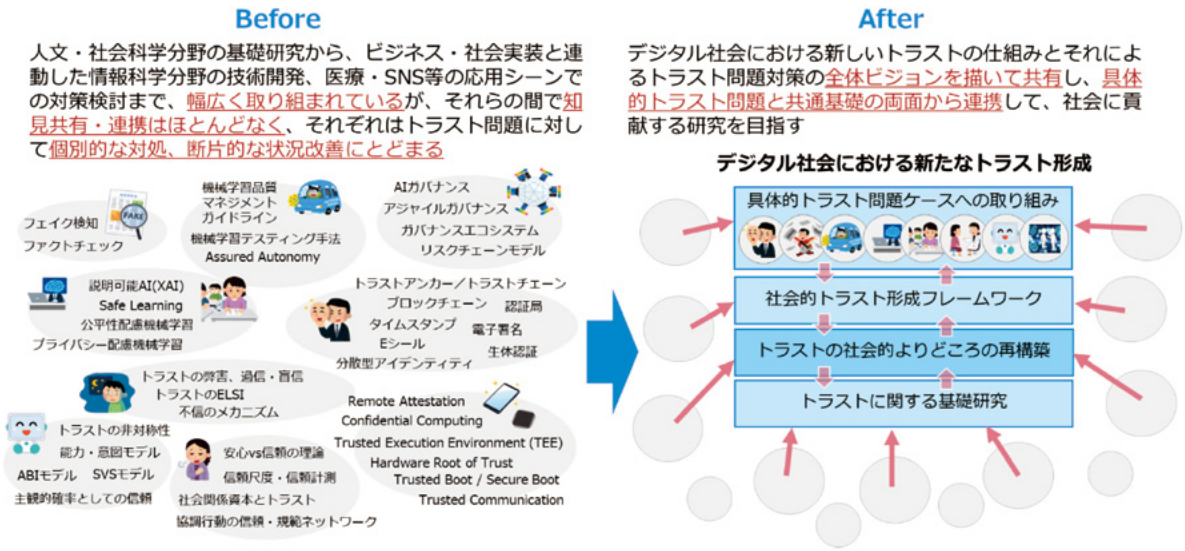


図1–4　　トラスト研究の状況と目指す姿（Before/After）

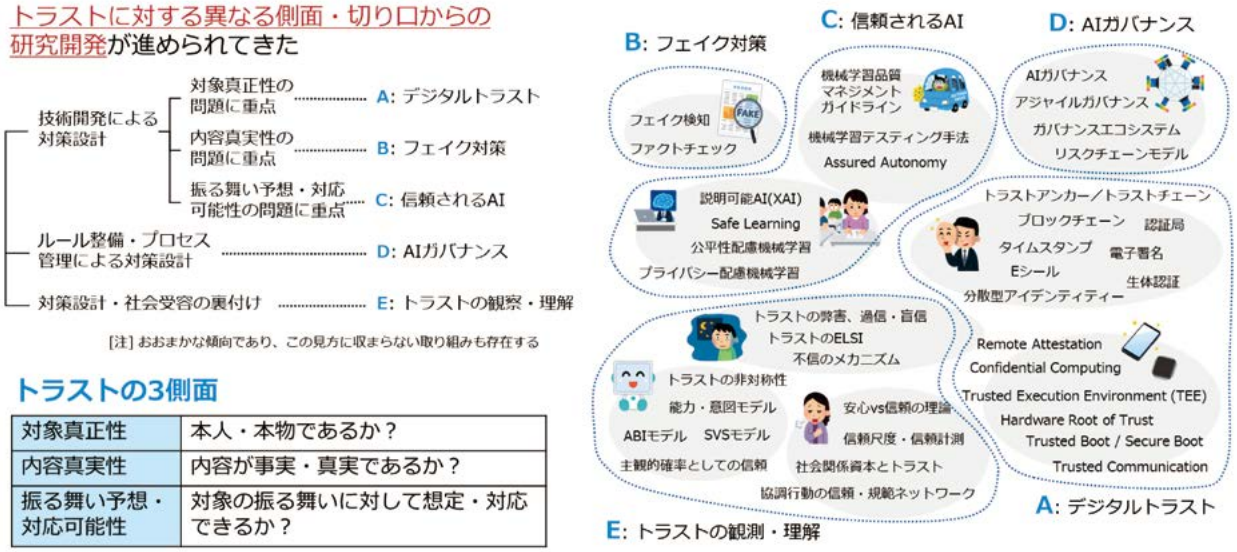


図1–5　　研究開発がばらばらに進んできた要因

そして、デジタルトラスト（A）の技術開発は対象真正性、フェイク対策（B）の技術開発は内容真実性、信頼されるAI（C）の技術開発は振る舞い予想・対応可能性にフォーカスし、技術開発による対策を考える取り組みである。これに対して、AIガバナンス（D）は、主にルール整備・プロセス管理による対策を考えている。また、トラストの観察・理解（E）に主眼を置いた人文・社会科学系の研究は、対策設計や社会受容の裏付けとなるものである。

このようなばらばらの取り組みでよいのだろうか。例えば、ある新しいサービスを使ってみようかと考えるとき、そのサービスがどのように動いてどのような結果が得られそうか、サービスの仕組みが信じられるか、というのは振る舞い予想・対応可能性であり、そのサービスの提供企業が怪しくないか、というのは対象真正性が関わる。さらに、そのサービスについての評判やレビュー投稿は本当か、ヤラセではないか、というのは内容真実性が関わる。われわれは、いろいろな視点から多面的に関連情報を集め、その一つだけでは確信を持てなくとも、それらを複合的に検証することで、総合的な判断を下すようなことをしている。

デジタル化の進展でリスクが高まった状況で、断片的に切り取られた情報や対象のある一面しか見ずに、何かを信じ込むことはとても危うい。デジタル社会のトラストは、多面的・複合的な検証によって支えていくべきであり、その検証が適切かつ容易に行えるような仕組みが望まれる。

表1–1　　主なトラスト関連政策提言・プログラム事例

分類	日本	海外
A. デジタルトラスト 対象真正性	● 「Data Free Flow with Trust (DFFT)」（2019年1月ダボス会議、安倍首相） ● デジタルトラスト協議会（2020年8月設立） ● 内閣官房デジタル市場競争本部 Trusted Web推進協議会「Trusted Webホワイトペーパー Ver.1」（2021年3月）、「同Ver.2（案）」（2022年7月） ● デジタルガバメント閣僚会議 データ戦略タスクフォース「包括的データ戦略」（2021年6月閣議決定） ● デジタル庁 データ推進戦略ワーキンググループ トラストを確保したDX推進サブワーキンググループ（2021年11月～）	● 欧州 eIDAS (electronic Identification and Authentication Service)規則（2014年7月成立、2016年7月施行）：トラストサービスの統一基準
B. フェイク対策 内容真実性	● CREST「信頼されるAIシステム」における「インフォデミックを克服するソーシャル情報基盤技術」（2020年度～）	● 米国国防高等研究計画局（DARPA）「Media Forensics (MediFor)」プログラム、「Semantic Forensics (SemaFor)」プログラム
C. 信頼されるAI 振る舞い予想・対応可能性	● 統合イノベーション戦略推進会議「AI戦略2019」（2019年6月）における主要な研究開発課題として「Trusted Quality AI」 ● 文部科学省2020年戦略目標「信頼されるAI」を受けたJSTプログラム：CREST「信頼されるAIシステム」、さきがけ「信頼されるAI」（2020年度～） ● NEDO「次世代人工知能・ロボット中核技術開発事業」において「AIの信頼性」（2020年度～）	● 英国研究・イノベーション機構（UKRI）「Trustworthy Autonomous Systems Programme」：自律システムの諸課題に対する学際研究、6つのノードとハブから成る体制で大学・産業界が多数参加
D. AIガバナンス	● 世界経済フォーラム「Rebuilding Trust and Governance: Towards DFFT」白書（2021年3月） ● 経済産業省「Governance Innovation Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」（2021年7月）、「AI原則実践のためのガバナンス・ガイドライン Ver. 1.1」（2022年1月） ● 日本ディープラーニング協会「AIガバナンスとその評価」研究会報告書「AIガバナンス・エコシステム－産業構造を考慮に入れたAIの信頼性確保に向けて－」（2021年7月）	● 欧州委員会「Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts」（2021年4月）

トラストに関連した政策提言やプログラムの事例を表1–1に示したが、トラストの各側面から施策が強化されつつあるのが分かる。ここでさらに、それらを包含する全体ビジョンを描いて取り組みを推進するならば、より強力な戦略構築や国際競争力の強化が可能になるのではないか。

研究開発の方向性

では、研究開発をどのように進めるべきか。それを説明する準備として、図1–6を用いてトラストのモデルを紹介する。

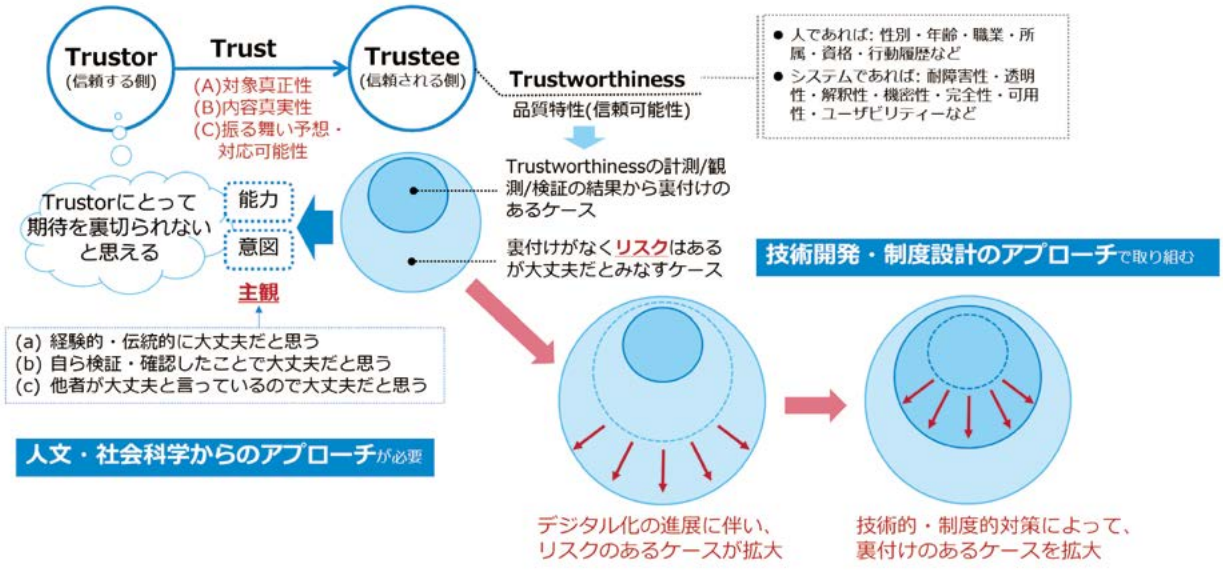


図1–6　　トラストのモデルとリスクを減らす方向性

Trustor（信頼する側）がTrustee（信頼される側）をトラストするかどうかというとき、図1–6では二重円で表現したが、期待を裏切らないかどうかについて、裏付けがある部分と裏付けがない部分がある。トラストは、このようなリスクがある状況でするものだということになる。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。リスクと主観はトラストを支える仕組みを実現していく際に、考えておかねばならないポイントだと思う。

この二重円の状況について考えると、デジタル化が進展して、リスクのあるケース、つまり外側の円が大きくなった。この状況を改善するため、技術開発や制度設計によって、裏付けのあるケース、つまり内側の円を拡大したいと考えるわけである。このように技術開発と制度設計が対策として重要なわけだが、最終的にどのような状況でトラストするかは、各Trustorの主観に依存する。この部分については、人文・社会科学の知見が有用である。したがって、分野横断の総合知による取り組みが必要である。

内側の円を拡大する、つまり、裏付けを持てるようにするためには、そのよりどころになるものが要る。先ほど述べたように、トラストするか否かは最終的にTrustorの主観に依存するとはいえ、詐欺のような犯罪を防ぎ、社会秩序を維持するため、社会的な共通認識となるようなよりどころが必要である。実際、現状でも、トラストの3側面のそれぞれに対して、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、社会的よりどころになっている。しかし、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ということなのだ（図1–7）。それゆえ、このトラストの社会的よりどころを拡充・強化・再構築することが必要になる。

トラストの3側面	現状の社会的よりどころ	問題点: トラストの綻び
(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・生体認証など	真正性保証の対象が拡大、デジタル特有の偽造・偽装・改ざんの可能性も拡大。信頼の基点の信頼性担保にも課題あり。
(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・監視カメラ映像など、学説は査読制による学術コミュニティ合意など	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な対象は限定的。
(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・ライセンスなど、機械・システムの動作については仕様書など	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も近似的説明であり、保証にはならない。

図1–7　　トラストの社会的よりどころの例と問題点

研究開発課題

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しい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として考えられるものを図1–8に示した。ここでは、研究開発課題を4層で捉えた。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1–9は、研究開発課題の第1層と第2層の実現イメージを、Before/Afterの対比で描いたものである。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

どころになるものが増強されている。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・複合的に検証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、分散監視のような機能も設けていくべきであろう。

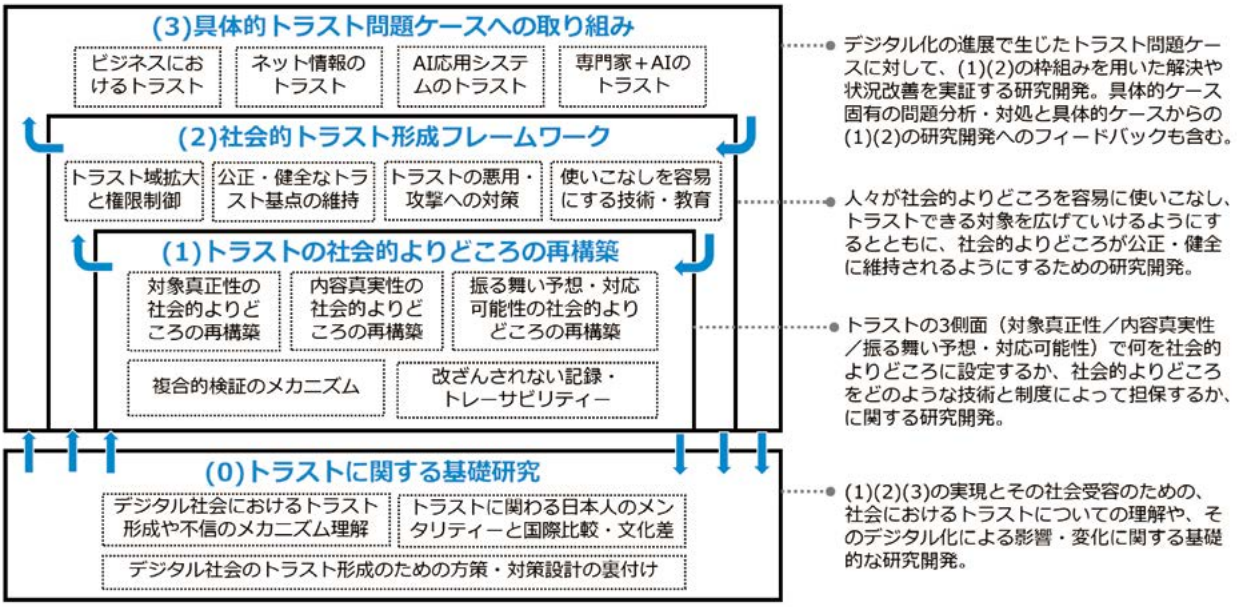


図1–8　　4層から成る研究開発課題

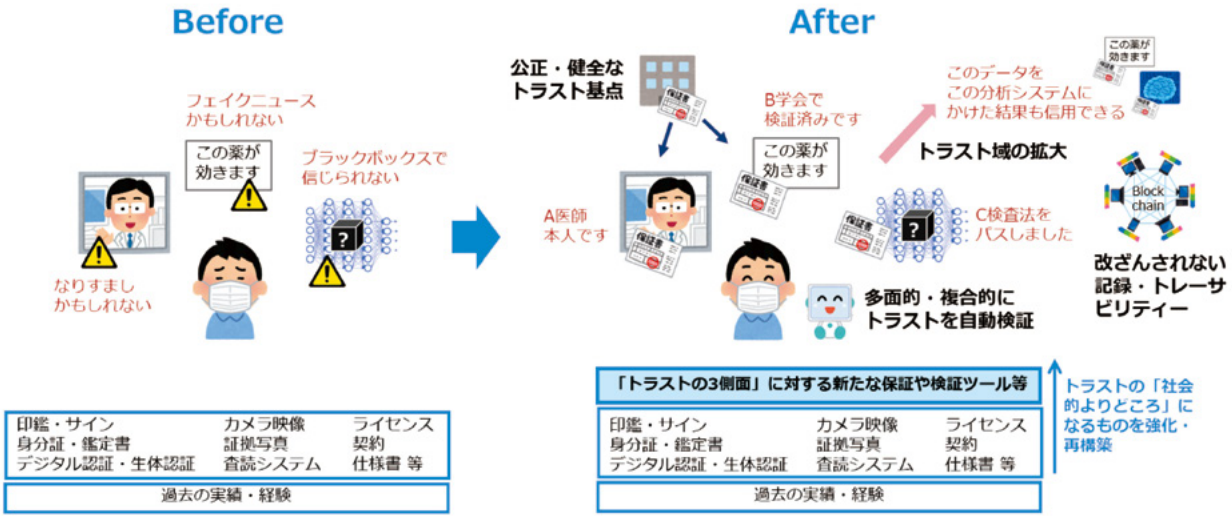


図1–9　　研究開発課題（1）（2）の実現イメージ

以上、2022年9月に公開した戦略プロポーザルから、問題意識や取り組みの方向性として考えてきたことを中心に紹介した。総合知による取り組み、幅広く分野横断的な議論が必要なテーマだと思っている。そこで、今回のシンポジウムは、キックオフ的な位置付けで、今後も継続して議論できるような場を設けていきた

いと考えている。

本日、この後は、第１部でトラスト３側面への取り組み事例について３件の講演、第２部でトラスト３側面の融合・分野横断の必要性について４件の講演、最後、第３部として総合討論という流れで進める。

2 ｜ トラスト3側面への取り組み事例

2.1［対象真正性］Trusted Web

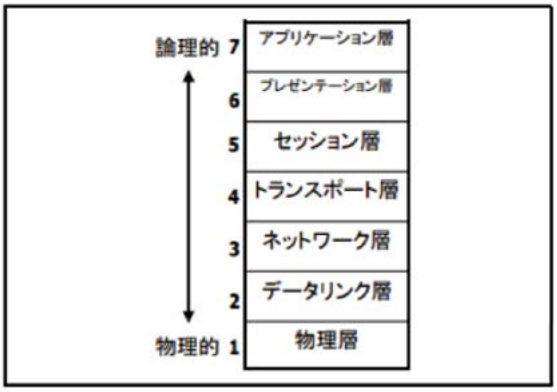
クロサカ タツヤ

Trusted Webについて説明する。Trusted Web推進協議会の委員および技術検討を進めるタスクフォースの座長を務めている。 Trusted Webは骨太戦略の中でも紹介されており、ホワイトペーパーのVer.2.0を2022年8に公開した。

2.1.1 インターネットにおけるトラスト

Trusted Webは、いわゆるWeb技術やWebサービスを対象としている。 Web技術やWebサービスは、Gmailなど、さまざまなアプリケーションの基盤となっている。では、このWebサービスおよびそれを支える基盤を含めた総体、つまりインターネットは、本当に信じられるのかということが大きな命題としてある。

インターネットをOSI参照モデルで表現すると、図2–1–1のように、第1層から第7層に区分され、それぞれ実装されている。物理層は、光ファイバーなどの物理的なメディアを使って情報を物理的に伝送するための層で、その上に、データリンク層とネットワーク層があり、一番上の第7層がアプリケーション層となっている。例えばIPアドレスは第3層のネットワーク層の識別子であり、端末をIPアドレスによって識別し、第4層のトランスポート層ではTCPやUDP、最近ではQUICというグーグルが提案した新しい第4層のプロトコルが実行されている。



出所：森下泰宏（JPNIC）「インターネットの基礎知識」Internet Week 99@パシフィコ横浜

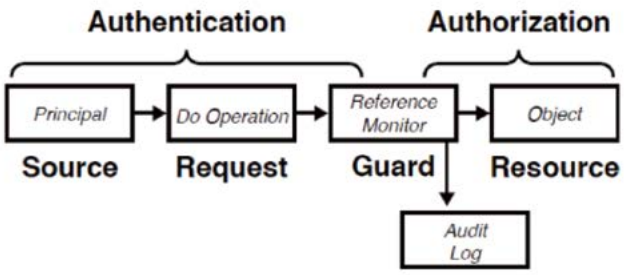
図2–1–1　　OSI 7層モデル

これらを組み合わせてインターネットが実現されているが、では第1層から第7層のどこでトラストを担保しているのかという点が重要な論点だが、現在のインターネットでは、どの層でもトラストを担保していないという状況にある。電子署名やタイムスタンプなどのトラストサービスは、第7層のアプリケーション層でトラストを担保しているが、それはインターネットのインフラの中の機能として標準化されたものではなく、エンド－エンドでお互いのやりとりを決めておいて、エンド－エンドでトラストを担保するという考え方になっている。トラストサービスが標準化され、社会制度の裏付けが保証されれば、第7層のトラストサービスでもトラストを担保できるという考え方もある。マイナンバーシステムなどは、この考え方を使って第7層でトラストを担保し

ている。一方で、第７層のアプリケーション層は一番提案がしやすいが、その反面、標準によって一本化するのが非常に難しい。しかしトラストが限定的あるいは暫定的な合意に基づくものであるとすると、そのトラストは、かなり脆弱なものとなってしまうと考えている。

そのため、Trusted Webでは、エンド－エンドでさまざまなトラストの仕組みを考えるよりも、インターネットの第７層以下の層、つまりインターネットのインフラの中にトラストの技術を埋め込んでトラストを実現した方が合理的ではないか考えている。もとよりインターネット全体で何らかの包括的にトラストのある状態を実現することは相当難しいのだが、そのチャレンジである。

図2–1–2に示すように、一般的なデジタルにおけるトラストでは、AuthenticationとAuthorizationがあり、その中にSource、Request、Guard、Resourceというプロセスがある。 Guardは、Requestしたエンティティーが対象のサービスを使う資格を持っているかどうかを判断し、持っていなければ却下する。 Auditは、その結果を記録する。Trusted Webでは、このように、Resource（Object）へのアクセスをGuardで制御することによってトラストを実現していこうとしている。この考え方は、従来のトラストの方法とも整合しており、例えば、電子署名の技術の一部をインターネットのインフラに埋め込むことでトラストを保証するということも可能となる。Trusted Webでは、この考え方に基づいて、トラストを保証するための基盤技術（ないしはミドルウェア）を作ろうとしている。



Source: Lampson, B. W.: Computer security in the real world, IEEE Computer, Vol. 37, No. 6, pp. 37–46, (online) DOI:10.1109/MC.2004.17 (2004).
https://www.cs.cornell.edu/courses/cs513/2005fa/NL02.Lampson.pdf

図2–1–2　　デジタルにおけるトラスト

2.1.2 Trusted Webにおける Trust

Trusted Webにおいてトラストがある状態というのは、どういう状態なのか、またトラストのレベルを設定するとしたらどういう度合いの問題として整理できるのか。トラストがある状態というのは「状態」なので、その状態は評価の仕方によって変わることになる。 Trusted Web協議会では、抽象的なレベルでトラストがある状態とは何なのかということを最初に３～４カ月かけて集中的に議論した。結論として、今のところ仮定しているのは、趣旨説明で紹介されたとおり、事実の確認しない状態で相手先が期待したとおりに振る舞うと信じる度合いと考えている。いちいち相手を確認したり検証したり、相手の振る舞いを検証しないと何もできないとすると、何かをやりとりするトランザクションが発生する都度、確認や検証が必要となり、効率が悪過ぎて使い物にならない。これに対して、都度、確認や検証するのではなく、相手を信じてみようというのがトラストのある状態だと考えると、そのためには必要を感じたときに相手を検証する、Verificationすることが必要となる。このVerificationの手段をできるだけ多く準備して、それらをインターネットのインフラに埋め込んでいこうというのが、Trusted Webで目指している姿だと考えていただければ良い。

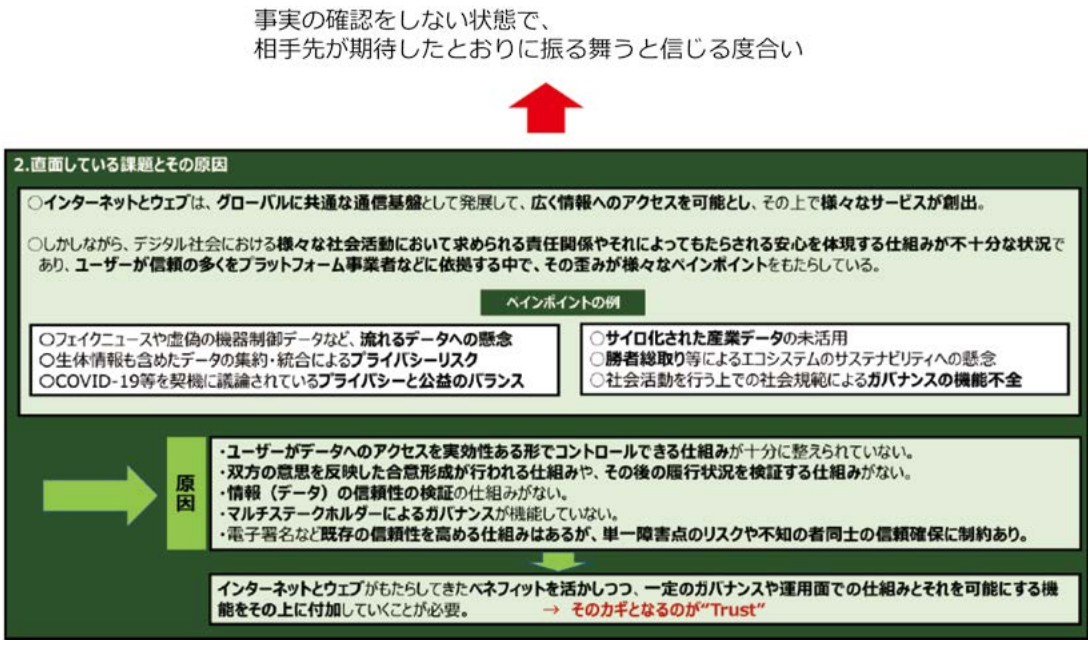
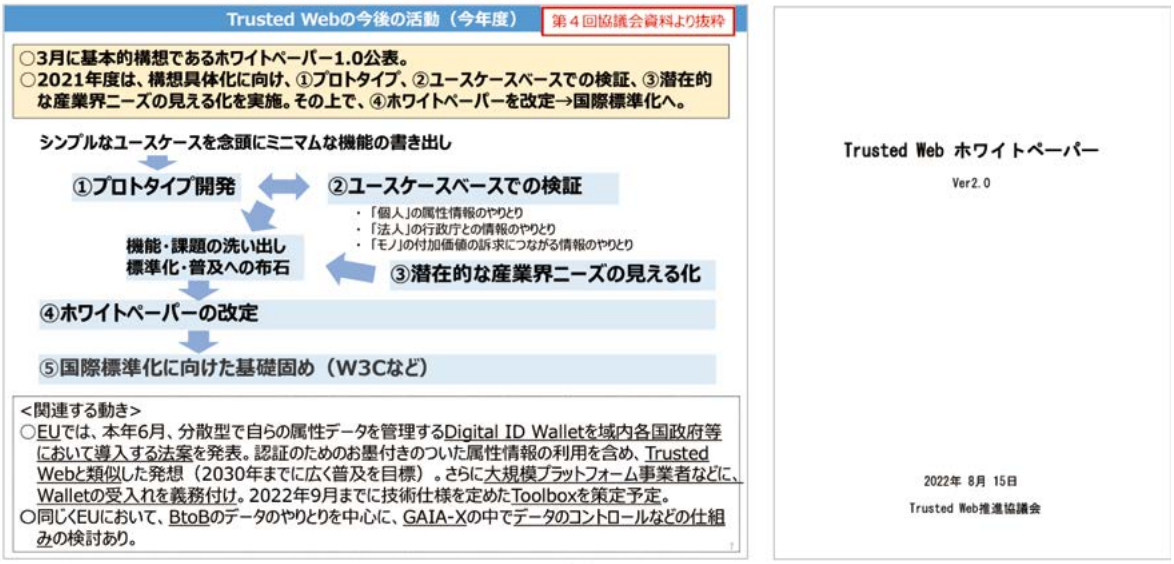


図 2–1–3　　Trusted Web における Trust



出所 https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/trustedweb.pdf

図 2–1–4　　Trusted Web の活動の概要

2.1.3 活動の概要

インターネットのインフラを作っていくということは、野望に近い話でもあり、いろいろな方と意見交換して合意形成をはかる必要がある。そのため、今は内閣官房の予算で活動しているが、最初から国だけではなく、産学官が協働で幅広く活動することを想定していた。国際標準化にも取り組むし、ユースケースからプロトタイピング、PoC（Proof of Concept）、そのためのソフトウェアの実装やサービス作り、インフラ作りの活動も始めている。これらの活動は、2022年8月に公開したホワイトペーパー（Ver.2.0）にまとめられている。以下では、ホワイトペーパーの内容の一部を紹介する。

2.1.4 直面している課題とその原因

まず、ペインポイントを具体的に定義している。例えば、フェイクニュースや虚偽の制御データなど、流れるデータへの懸念、生体情報も含めたデータの集約・統合によるプライバシーリスク、プライバシーと公益をどうバランスしていくのか、企業活動での営業秘密の観点から、どうしても相手にデータを渡したくないというサイロ化された産業データの問題、プラットフォーマーの勝者総取りによるエコシステムのサステナビリティーの懸念、それによりガバナンスが偏ったものになるのではないかという問題などが起きるのではないかと考えている。この状態を、完全に解決できるとはいえないが、少しでも緩和していくためにトラストをインフラに埋め込んでいく必要があるのではないかと考えている。

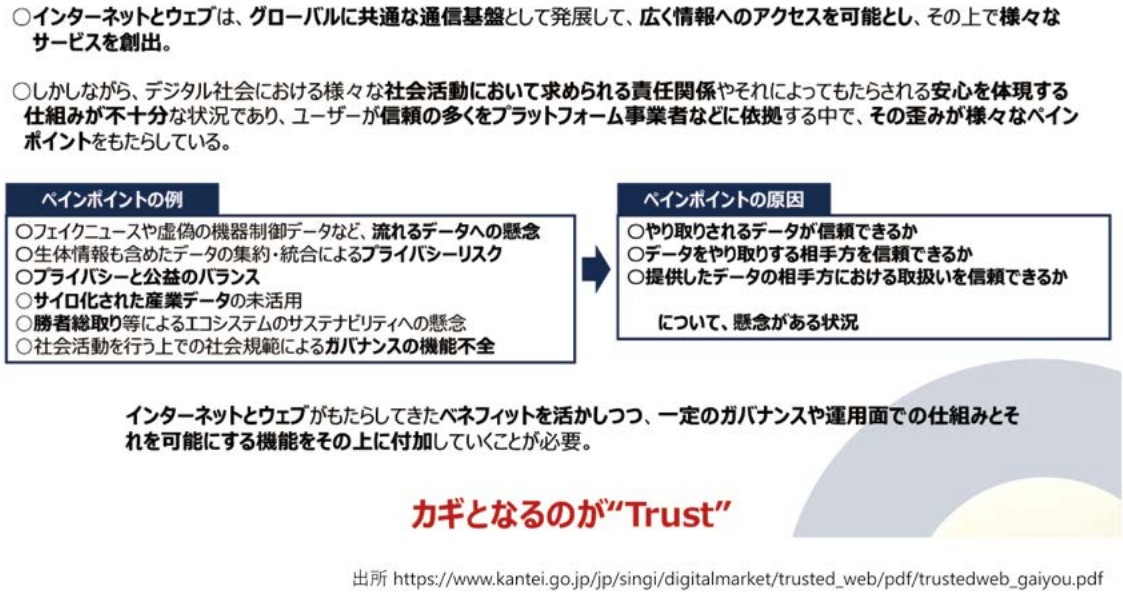
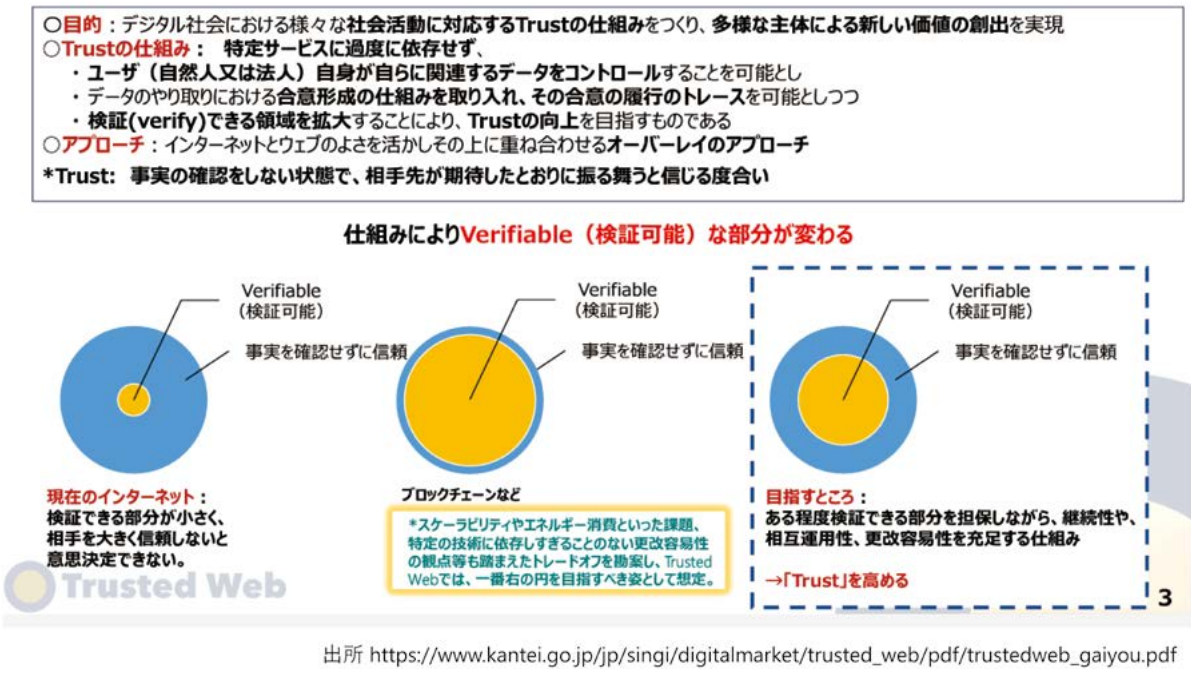


図2–1–5　　直面している課題とその原因

2.1.5 Trusted Webが目指すべき方向性

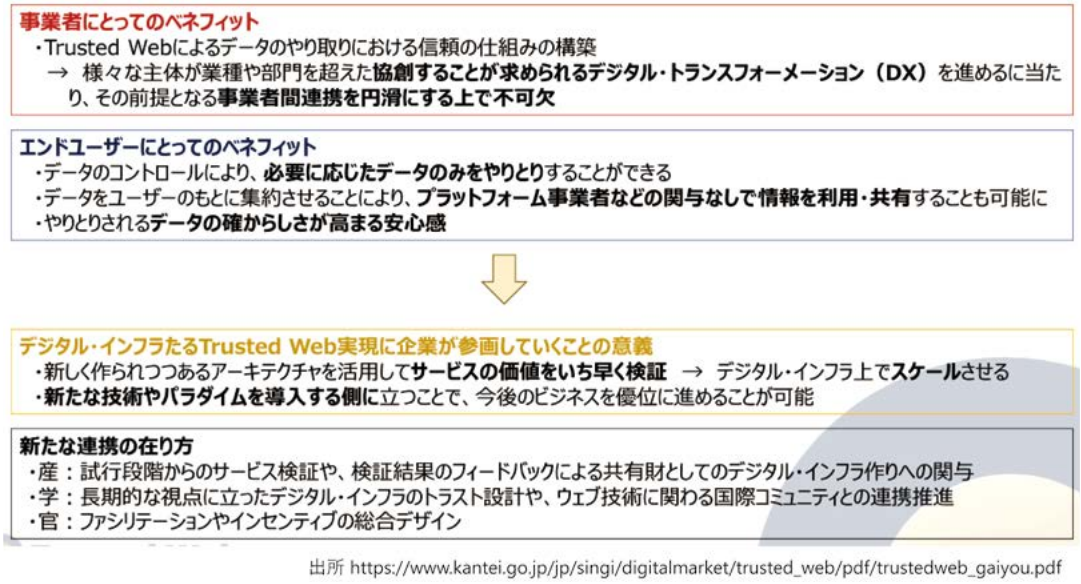
今のインターネットでトラストを担保する考え方としては、二つのパラダイムがある。一つは、プラットフォーム事業者に全てを委ねて、プラットフォーム事業者がトラストを担保するという考え方である。例えば、グーグルがアイデンティティマネジメントなどを行うことによりトラストを担保する。もう一つは、プラットフォーム事業者に全てを委ねるのではなく、ブロックチェーンによりトラストを担保するという考え方である。現在、この二つの考え方があるが、どちらも都合がいいわけではない。恐らく、誰もプラットフォーム事業者に全てを委ねるのがいいとは思っていないし、ブロックチェーンにも問題がある。ブロックチェーンは、いつでも検証できる、常に検証するということをプルーフ（証明）することによりトラストを担保しているが、このプルーフのためのコストが非常に大きい。そのため、ブロックチェーンでは、大量の情報を扱うことが難しく、トークンのような小さい情報しか扱うことができない。この場合、そのトークンが本当に正しいのかという問題が出てくる。このように、ブロックチェーンも完全ではなく、いつでも検証できる、常に検証するだけでは、不十分である。Trusted Webでは、1回信じてみよう、だけど、必要があれば検証できる状態を既存のインターネットより増やそうということを理念としている。つまり、ブロックチェーンの考え方が「Don't Trust, Verify（信じるな、検証せよ）」だとすると、Trusted Webの考え方は「Trust, but Verify（信じよう、だけど、必要

なときには検証しよう）」といえる。



出所 https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/trustedweb_gaiyou.pdf

図 2–1–6　　Trusted Web が目指すべき方向性



出所 https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/trustedweb_gaiyou.pdf

図 2–1–7　　Trusted Web のもたらすベネフィット

2.1.6 Trusted Web のもたらすベネフィットとそのユースケース

Trusted Web には、いろいろなベネフィットがあると考えている。これに関しては、企業や事業者からいくつかの話を聞いている。例えば、企業でサプライチェーンの効率化を考えるときに、全てをプラットフォーム事業者に頼むということは受け入れがたい。一方で、全てをブロックチェーンで実現できるかというと、実際には難しい。このため、自分たちのサプライチェーンのために何かミドルウェア的なものを作らないといけな

いが、どのようなものを作ればいいのか悩んでいたという意見が多数あった。また、企業だけでなく行政サービスやサードパーティーなどいろいろな企業が関わるトランザクションを考えるときにも、何かを作らないといけないが、どう作ればいいのか分からないという意見が結構あった。

このため、Trusted Webの仕組みが共通基盤になるのであれば、事業者やエンドユーザーにとってもベネフィットがあるという声をいろいろな所から聞いている。ベネフィットをさらに洗練して、実際にインフラやミドルウェアなどの共通基盤に近い部分を作っていこうと考えている。

具体的なユースケースの分析やプロトタイピングを進めていく中で、価値創造につながることが期待されるケースが少しずつ見えてきている。サプライチェーンや相互評価のトラスト、トランザクションがとにかく多く確認コストが高い分野、個人（法人）によるコントロールのニーズが高い分野、大量のIDやデータを持っていながら、さらなる活用が考えられる分野などが挙げられる。

個人、法人、サプライチェーンという三つのユースケースについて具体的な検討を進めるとともに、「個人」の属性情報のやりとりについては、プロトタイプで実際に動かしてみた。その結果の詳細については、ホワイトペーパーに記載があるので参照していただきたい。また、2022年の秋から新しいユースケースの実証を始めており、11事業者からの提案を採択した。これらの提案では、個人、法人、サプライチェーンや、個人の中でも機微性がある領域や複雑な領域を扱っており、実際にプロトタイプを開発しているという状況である。この取り組みの中で、共通基盤をどう作るのか、その上でアプリケーションはどのように作れるのかということの議論も進めている。

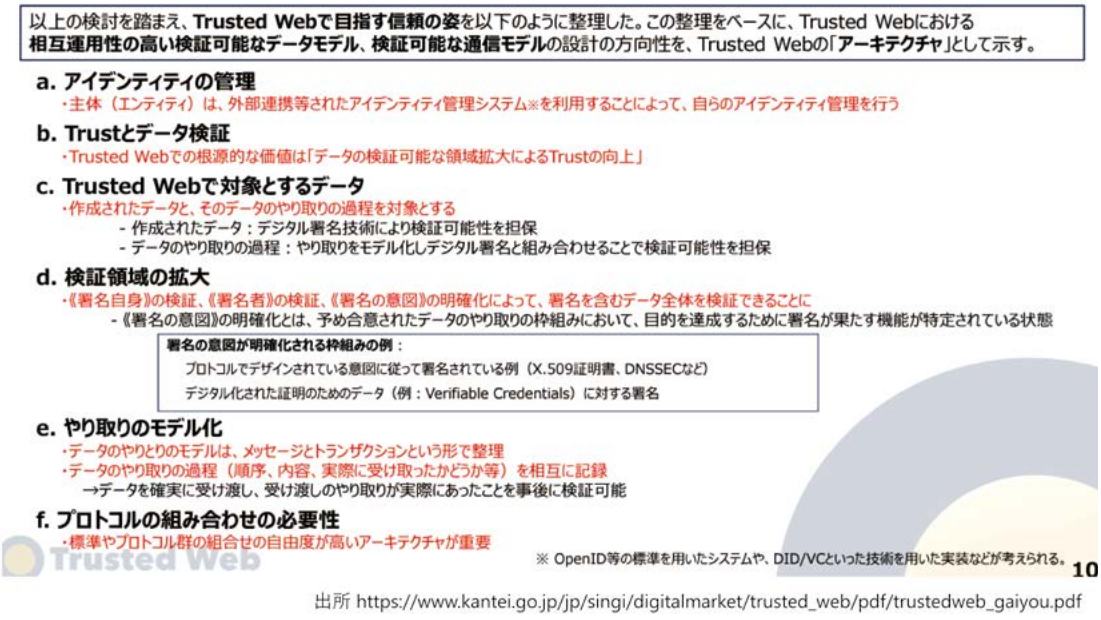


図2–1–8　　Trusted Webで目指す信頼の姿

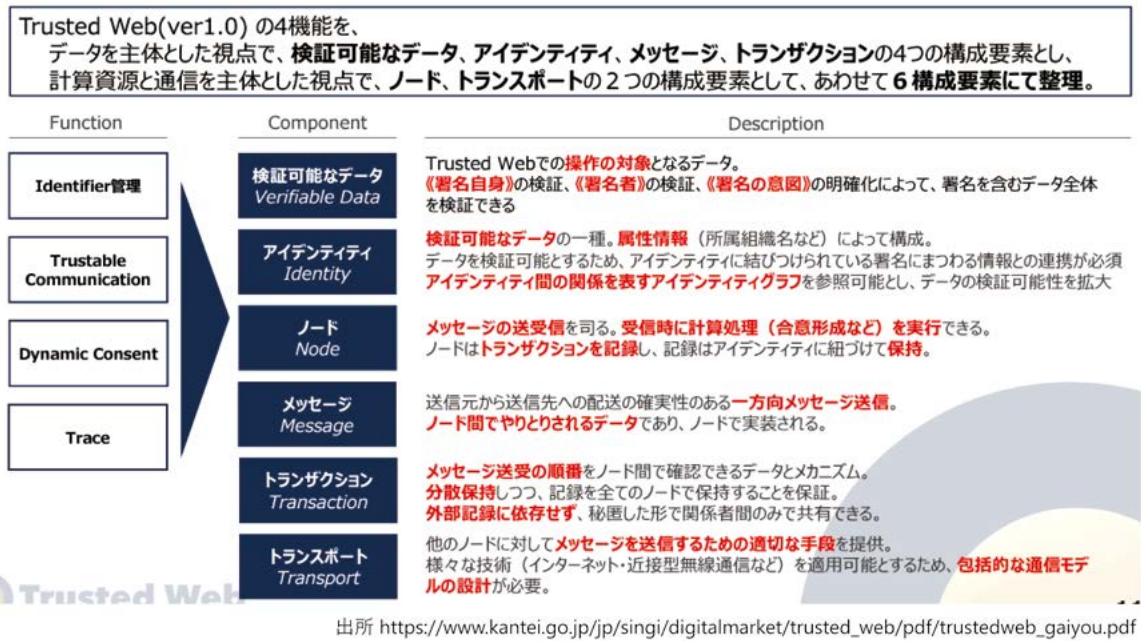
2.1.7 Trusted Webで目指す信頼の姿とそのためのアーキテクチャー

Trusted Webの目指す姿は、先に説明したように、簡単に言えば「目をつぶってでも、1回、信じてみよう、だけど、怪しいなと思ったら確かめられるような状態を作ろう」ということである。

そのためには、アイデンティティーの管理やデータがどのような状態にあるのかを検証できるようにしておくこと、対象とするデータが一体何なのかを明確にすること、検証領域をできるだけ広げるということが重要で、これらをどう実現するかや、やりとり（トランザクション）のモデル化をきちんとしていくこと、どこか特定の

レイヤーにプロトコルを埋め込めばいいということではないので、プロトコルの組み合わせの必要性を考えること、これらのことをシステムを作るという観点で、今、議論しながら整理している。

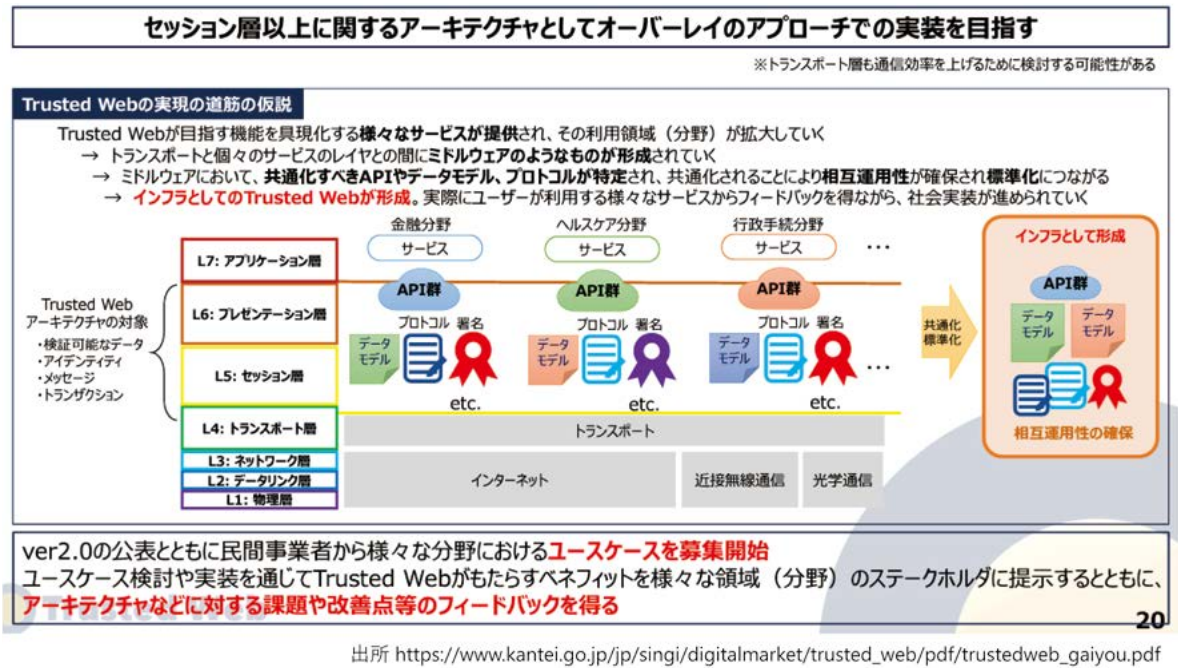
Ver.1.0の四つの機能（Identifier管理、Trustable Communication、Dynamic Consent、Trace）を、実装のための六つのコンポーネントとして整理している。この六つのコンポーネントがTrusted Webの構成要素であり、四つの機能と六つのコンポーネントで構成されていればTrusted Webといっていいのではないかと議論している。現在議論している内容を踏まえて、ホワイトペーパーVer.3.0が2023年春から夏に出ると思う。



出所 https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/trustedweb_gaiyou.pdf

図2–1–9　　Ver.1.0での４機能を６構成要素にて再整理

Trusted Webのアーキテクチャーは、レイヤー構造になっている。上位層にエンティティー、下位層にノード、その間にデータモデルがある。その中でもアイデンティティーマネジメント技術は重要なコンポーネントである。アイデンティティーは基本的に人間や、システムを使うエンティティーを特定し、アイデンティティーマネジメント技術はそれらを管理する。ユースケース分析を通して、実際にはエンドユーザーだけではなく、トランザクションやその関係性も含めて管理する必要があることが分かってきた。例えば、サプライチェーンでは、単に企業A、企業B、企業Cが特定できればいいだけでなく、企業Bがこういう状態だから企業Aから企業Cへのトランザクションを仲介してよいとか、企業Bがこの状態でなかったら企業Aから企業Cへのトランザクションを仲介してはいけない、バイパスしないといけないとか、途中で止めなければいけないとか、などさまざまな状況があり、これらを考慮する必要がある。このように、エンティティーは、実際のトランザクションやデータの状態に応じて制御が変わる。この制御を、どのレイヤーで、どう実現していくかを検討するために、構成要素を抽象化して整理していこうとしている。また、コンポーネントの組み合わせについても整理しようとしている。



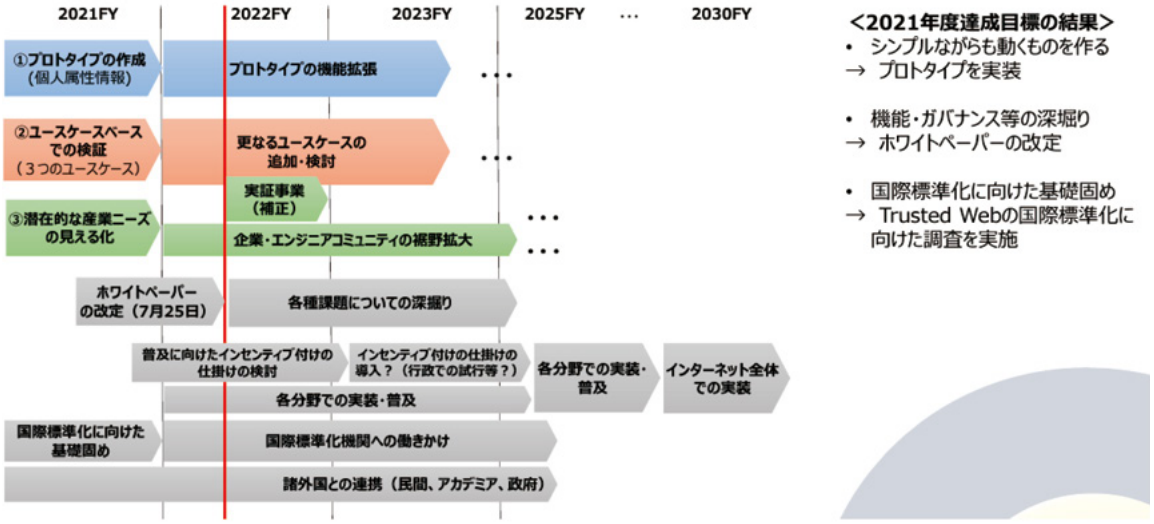
出所 https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/trustedweb_gaiyou.pdf

図2–1–10　　オーバーレイの考え方と実現に向けた道筋

オーバーレイとは、あるレイヤーの上に新しい機能をかぶせる方法で、現実的なアプローチだと考えている。どのレイヤーにオーバーレイするのかということは議論が必要だが、あるレイヤーを全て作り直すというわけではないので、新しい機能を提案する場合、オーバーレイを採用した方が実装や普及が早いと考えられる。このため、Trusted Webではオーバーレイのアプローチによる実現を目指している。

2030年頃に、今回説明した内容が全てできあがっているという想定で取り組みを進めている。長い話ではあるが、合意形成に時間がかかる話でもあるので、着実に進めていきたいと考えている。

以上、Trusted Webでのトラストの考え方やシステム作り、合意形成をしようとしているといった点などを紹介した。



出所 https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/trustedweb_gaiyou.pdf

図2–1–11　　2030年に向けた中期的な戦略（イメージ）

2.2［内容真実性］メディアのトラスト

山口 真一

「デジタル社会におけるメディアとトラスト：偽・誤情報の状況から」というタイトルでお話しする。

私は経済学博士で、特に計量経済学というデータ分析手法を使ってソーシャルメディア上のフェイクニュース、ネット炎上、誹謗中傷といった諸課題、あるいは情報社会の新しいビジネスモデル、経済法則について研究している。3年以上グーグルと一緒にフェイクニュースの研究をしており、総務省とは昨年度、偽・誤情報に関する啓発教育教材[2]を作ったり、さまざまな委員会に参加したりもしている。

2.2.1 世界を脅かす偽・誤情報

インターネット、ソーシャルメディアが普及し、誰もが自由に世界へ情報発信できるようになった。以前は著名人やマスメディアしかできなかったことだ。まさに革命的な出来事で、人類総メディア時代が到来したと言える。

こうした自由な情報の流通が経済・社会システムを大きく変えつつある。ポジティブな面では、われわれは時間や地域の制約なしに世界中の誰とでもコミュニケーションをとることが容易になった。それに伴ってマイノリティーの人が同じ境遇の人とつながりやすくなった。

グーグルとのInnovation Nipponというプロジェクト内で以前、ネット上の口コミの経済効果を分析した結果、ネット上の口コミによって人々の消費が年間およそ1.5兆円押し上がっていることが分かった[3]。このようにソーシャルメディアは社会、経済、さまざまな面でポジティブなインパクトをもたらしている。

しかし同時に、大きなネガティブなインパクトも目立ってきた。2016年は偽・誤情報元年、フェイクニュース元年と言われる。2016年の米国大統領選挙では偽・誤情報がさまざまに飛び交った。例えば「ローマ法王トランプ氏を支持　世界に衝撃」という偽・誤情報があった。あるいは「ヒラリー・クリントン候補が児童買春組織に関わっている」という偽・誤情報を信じた人が、その拠点とされたピザレストランで発砲してしまった事件。この選挙前の3カ月間でトランプ氏に有利な偽・誤情報は3000万回、クリントン氏に有利な偽・誤情報は760万回もシェアされたことが分かっている。その後も、選挙、EU離脱の国民投票といったイベントのある度にこのような偽・誤情報が多く拡散することが確認されている。

偽・誤情報は選挙や政治に限らない。例えばメキシコの事件だが、路上飲酒の男性2人が警察署で事情聴取を受けていたところ「彼らが実は子どもの誘拐に関わっていた」とのデマがメッセージアプリ上で拡散した。その結果、2人は釈放されるやいなや集団リンチを受けて殺害されてしまった。真っ赤なうそを基に殺されたのだ。

2.2.2 災害・感染症はデマが拡散しやすい

日本は災害大国と言われるが、災害に際してさまざまな偽・誤情報が流れているという現実もある。有名なものでは、2016年熊本地震の際に「動物園からライオンが逃げた」というようなツイートをした人がいた。それが2万件弱もリツイートされ、動物園にも市にも問い合わせが殺到。災害直後の対応で非常に大変な時期に、別の業務が大量に発生した。この投稿をした人は偽計業務妨害で逮捕されている。

2 総務省「【啓発教育教材】インターネットとの向き合い方～ニセ・誤情報に騙されないために～」
https://www.soumu.go.jp/use_the_internet_wisely/special/nisegojouhou/

3 Yamaguchi, S., Sakaguchi, H., & Iyanaga, K. (2018). The Boosting Effect of E-WOM on Macro-level Consumption: A Cross-Industry Empirical Analysis in Japan. The Review of Socionetwork Strategies, 12(2), 167-181.

2018年北海道地震が発生した後、「さらに大きな地震がくるぞ」というような偽・誤情報がメッセージアプリ上で拡散。それに対して自治体が「それは根拠のない情報ですよ」とアナウンスせざるを得なかったという事例もある。

2022年台風15号による静岡県の災害の際にはツイッター上に「ドローンで撮影された静岡県の水害。マジで悲惨過ぎる…」という投稿。実はAIで創作された画像だったのだが、これを投稿したことによって実際に信じた人がいて、拡散されていった。

ディープフェイクが民主化してきているのがよく分かる事例だ。AIを使った画像生成は、昔であればかなり高度な技術を持っている人が、悪意やいたずら目的でやったと思う。ところがこの投稿者はAIのスペシャリストでも何でもない。今は画像生成AIのサービスにアクセスして文字を入力するだけで、そのテキストに沿ったディープフェイク画像を簡単に作れてしまう。

現在、新型コロナウイルスが世界で依然として猛威を振るっている。コロナに関してもさまざまなデマや陰謀論が大量に拡散され、WHOはこれをインフォデミックと呼んで警鐘を鳴らした[4]。例えば「アルコールを飲めば予防、あるいはウイルスを死滅させられる」といったデマが世界中で拡散し、それを信じた人たちがメタノールを飲んで中毒になり、イランでは800人以上が死亡した。また「5Gの電波が新型コロナウイルスを運んでいる」との偽・誤情報を信じた人が、実際に電波塔を破壊しに行った事例もある。

災害や感染症の際、非常にデマが拡散しやすいことは歴史的に見ても明らかになっている。その背景として、参考になりそうな統計データを紹介したい。グーグルと行っているInnovation Nipponプロジェクトで、コロナワクチンに関するデマの拡散について分析したものだ。コロナワクチンデマを拡散したときに、その情報を信じていた人と、分からない・どちらともいえないと判断を保留していた人たちがそれぞれ、なぜその情報を拡散したかという理由の回答結果である[5]。

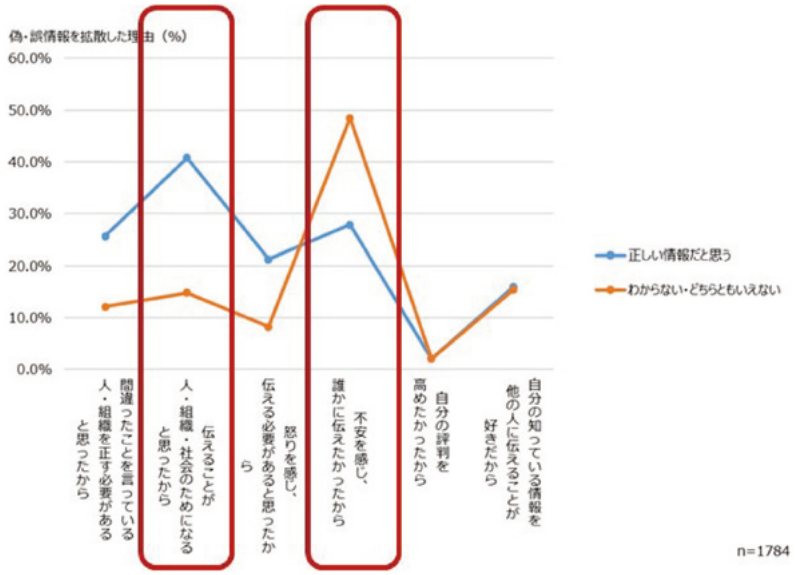


図2–2–1　　コロナワクチンデマを拡散した理由

4 WHO “Coronavirus disease (COVID-19) advice for the public: Mythbusters” https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters

5 山口ほか（2022）Innovation Nippon 2021「わが国における偽・誤情報の実態の把握と社会的対処の検討　―政治・コロナワクチン等の偽・誤情報の実証分析―」https://www.glocom.ac.jp/activities/project/7759

真偽の判断を保留している人たちの中で最も多かった拡散理由は「不安に感じたから」。正しい情報だと思っている人が拡散した理由で最も多かったのは「伝えることが人・組織・社会のためになると思ったから」である。災害、感染症はまさに不安に感じたり、これを早く人に伝えなければと感じたりするものを多く発生しやすいジャンルなので、デマが拡散しやすい。

現在ロシアがウクライナに侵攻しているが、この戦争においても偽情報が非常に重要な役割を果たしている。「ブチャで虐殺があった」との情報に対して、ウクライナ側は「それはロシア軍がやった」というふうに主張、ロシア側は「それはウクライナがやった」というふうに主張した。その際には衛星写真を使ったファクトチェックによって、ロシア軍が撤退する前に既に遺体が転がっていたことを突き止めたことで、ロシアの主張は否定された。また、ゼレンスキー大統領が降伏を呼びかけているかのようなディープフェイクの動画も拡散された。

高度な技術を使ったファクトチェック、あるいはディープフェイク動画の情報を事前にキャッチしていて、ウクライナはかなり早期に対応したので、すぐ収束した。しかしそうした準備ができていなかったら、ものすごく拡散されて大混乱していた可能性もある。

偽・誤情報は日本も対岸の火事ではない。日本でも年間2600件の疑義言説が拡散していることが分かっている[6]。2020年の米国大統領選挙では「トランプ氏による選挙不正」という主張が日本でも拡散され、デモを起こす人々もいたほどの大きな動きがあった。ある偽・誤情報については、英語圏よりも日本語圏のほうが拡散されたというデータもある。

2.2.3 日本における偽・誤情報の実態

偽・誤情報の実態を調査するために、コロナワクチン関連六つと政治関連六つ、2分野で計12個、実際にあった偽・誤情報を使ってアンケートを行った[7]。

まず、偽・誤情報への接触率。こういった偽・誤情報を、知っているという人が2%から25%ぐらいまである。また、12個のうち一つ以上接触していた人が40.4%であった。拡散量が多かった偽・誤情報を選んであるとはいえ、たった12個のうちいずれかのフェイクニュースを、5人に2人が知っている。非常に多くの人が偽・誤情報に触れていることがこの結果から分かる。

これは極めて重要なことだが、コロナワクチン関連の偽・誤情報に接触したときに、それが誤っていると気づいた人が平均で約43.4%しかいないことが分かった。残りの人は信じている、正しい情報だと思っているか、分からない、どちらとも言えないと判断を保留している。政治関連の偽・誤情報ではこれが何と20.3%になる。つまり、政治のフェイクニュースに接触した人のうち、5人に4人はそれが間違っていると気づかずに生活している。これらはいずれもファクトチェック済みの情報にもかかわらず、実際には大半の人が誤りだと気づいていない。

こうした偽・誤情報はネットを多く使っている若者だけの問題と捉えがちだが、年齢は、実はあまり関係ない。しかも政治的な偽・誤情報となると、むしろ50代、60代の方がよほど気づきにくい傾向も見られる。どの年代の人も偽・誤情報に気をつけましょうと言いたい。

残念ながら、偽・誤情報の拡散に関わっている人は少なくない。平均すると接触者の約15～35%は拡散している。注目すべきは拡散手段だ。突出して多いのは「家族、友人、知人に直接の会話で伝えた」。つまり、SNSやメッセージアプリではなく会話で、最も偽・誤情報を拡散している。インターネット上で厳格に規制す

6 シエンプレ デジタル・クライシス総合研究所「デジタル・クライシス白書2020」

7 国際大学グローバル・コミュニケーション・センター 2022年4月 Innovation Nippon 報告書「わが国における偽・誤情報の実態の把握と社会的対処の検討 ―政治・コロナワクチン等の偽・誤情報の実証分析―」 http://www.innovation-nippon.jp/reports/2021IN_report_full.pdf

ればよいという話でもなく、もっと社会全体で抜本的に考えていくべき問題である。 SNSだけに気をつければよいわけでなく、家族、友人、知人の話でもしっかりと情報を検証していくことが大切だ。

2.2.4 フェイクが広まる理由、作られる理由

なぜ、偽・誤情報がこんなに広まるのか。その拡散メカニズムについて、モデルを使い定量的に分析した。

まず、信じている情報は非常に拡散しやすい。ところが、誤っていると気づいている人はそれを拡散しない傾向にある。つまり、こんな情報があったよという拡散は多いが、この情報は間違っているよという訂正情報は出にくいのだ。さらに、メディアリテラシーや情報リテラシーが低い人ほど拡散しやすい。

偽・誤情報を信じている人、リテラシーの低い人が情報を拡散しているのが、今、われわれが接している言論空間だと言える。

実際、真実が1500人以上に到達するにはフェイクニュースより約6倍の時間がかかるとの研究結果がある。偽・誤情報はものすごいスピードで拡散していく一方で、真実は拡散スピードが遅く全然広まらない[8]。

なぜ、偽・誤情報がこんなに作られるのか。その背景は二つある。

一つは経済とアテンションエコノミー。現代は人々の注目をページビューに、それを広告収入につなげることが非常に強化されている時代だ。しかも、ぱっと見て「何だろう、これ」と思わせる、システム1と言われるような速い思考による注目、アテンションエコノミーの中では、センセーショナルなタイトルをつけたフェイクニュースを出し、それが見られることによって大きな広告収入を得るというメカニズムが働いてしまう。例えば2016年の米国大統領選挙では、マケドニア共和国という小さな国の学生が大量の偽・誤情報を作成し、親の生涯年収を超えた。フェイクニュースを作ることが一つの産業になってしまっている。

もう一つは政治的動機。自分の支持政党に対立する人たちの偽・誤情報を流すことはかなり行われている。2016年の米国大統領選挙では、ロシアからの介入があったと指摘されているし、外国だけでなく国内からも偽・誤情報はしばしば出る。

2.2.5 法規制には大きなリスク

では、偽・誤情報にどう対処していけばよいか。

まず私の調査では、「フェイクニュースには法規制が必要である」と答えた人が実は日本全体で74%もいた。フェイクニュースに法規制をすべきであると、4人に3人が考えている。

しかし法律にはリスクがある。例えばマレーシアでは偽ニュース対策法が作られたが、当時の政権が自分の敵となる人を捜査するのにそれを悪用した。その結果、こういう法律は良くないとして、政権交代後に廃止することになった[9]。ロシアではウクライナ侵攻に伴い、フェイクニュースを規制する法律が強化されたが、それも戦争に反対する人などを次々と逮捕するために使われてしまっているのが現状だ[10]。

つまり、偽・誤情報は非常に定義が難しい。トランプ元大統領も、マスメディアからは彼がフェイクニュー

8 山口真一（2022）『ソーシャルメディア解体全書』、勁草書房
Vosoughi, S., Roy, D.K., & Aral, S. (2018). The spread of true and false news online. Science, 359, 1146-1151. https://doi.org/10.1126/science.aap9559

9 日本経済新聞 2019年10月9日「マレーシア、偽ニュース対策法廃止へ　下院が再び可決」
https://www.nikkei.com/article/DGXMZO50821000Z01C19A0000000/

10 朝日新聞デジタル 2022年3月5日「ロシア、フェイクニュースと見なせば禁錮刑に　欧米メディア取材停止」
https://www.asahi.com/articles/ASQ356J15Q35IIPE00N.html

スを言っていると言われ、彼はマスメディアがフェイクニュースを言っていると言っていた。偽・誤情報かどうかは結局、人によってかなり判断が変わってしまう、固定されていないもの。それに規制を敷いてしまうと、やがて表現規制につながり、逮捕のために悪用されるという大きなリスクがある。

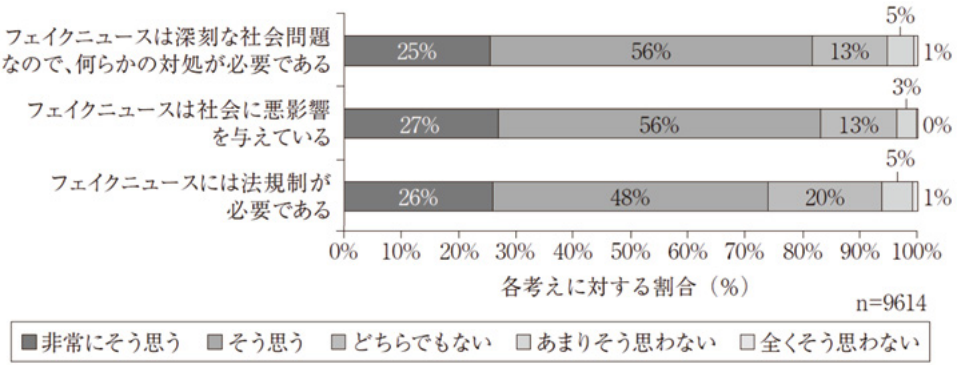


図2-2-2　　フェイクニュースについてどう考えているか

山口真一（2022）『ソーシャルメディア解体全書』、勁草書房

プラットフォーム事業者に削除義務を課すという対処もある。ドイツではネットワーク執行法というものがあり、侮辱などの違法な内容があるとユーザーから報告された場合、直ちに違法性を審査して24時間以内に削除する必要がある。この対応が不十分な場合、最高5000万ユーロまでの過料が科される。

これに対して総務省は「こういった法的規制が導入検討されているが、表現の自由の萎縮効果への懸念の声があるほか、立法後に違憲判決により当該規定が削除されるといった状況があった」との見解を示した。これはフランスの事例を指している。フランスでは似たような法律が出て、表現の自由の観点から違憲であるとされたのだ[11]。

したがって総務省の見解は、できるだけ自主的スキームを尊重し、法規制には極めて慎重な検討を要するということで、私も非常に賛成だ。ただ、ドイツのネットワーク執行法に関して、プラットフォーム事業者が罰金から逃れるため過剰に削除してしまう、いわゆるオーバーブロッキングの問題は、今のところは発見されていないと聞く。こうした法律がどのような影響を及ぼすかは、中長期的な観点での精査が求められる。

2.2.6 プラットフォーム事業者の透明性とさまざまなデマ対策

では、どうすればよいか。総務省が発表している資料では、「透明性、アカウンタビリティーの確保を求めていくべきではないか」と述べた[12]。例えばプラットフォーム事業者が違法/有害情報対策として何をしているか、実際にどれぐらい削除したか、その企業が日本語でしっかりとユーザーの相談に乗れる体制を作ってあるか、また取り組みの効果について、分析して公開することが求められると。私も総務省と同じように考えている。

最近こうした「透明性レポート」が世界的に発表されているが、日本国内のデータをぜひ積極的に公開していただきたい。やっている事業者も最近出てきたが、まだまだ不十分な点もある。ユーザーからの要望に日本語で丁寧に対応できる体制の構築も大切だ。

11　総務省2022年6月「プラットフォームサービスに関する研究会　第二次とりまとめ（案）」
https://www.soumu.go.jp/main_content/000823588.pdf

12　総務省2022年6月「プラットフォームサービスに関する研究会　第二次とりまとめ（案）」
https://www.soumu.go.jp/main_content/000823588.pdf

透明性に加えて、もう一つプラットフォーム事業者に求められるのが、アーキテクチャーによる対策の促進だ。昨今、偽・誤情報に対するラベルづけ、あるいは読まずに拡散しようとすると「それを本当に拡散するのですか」という警告が出るなどが仕様になっている。既存の取り組みを推進していくとともに、さらなるサービス改善を検討、実装していくことが大事なのではないか。ファクトチェック結果が優先的に届く仕組みの構築も求められる。

フェイスブックの事例では、フェイクニュースに対する対策を採ったらエンゲージメント数が半分以下になったという[13]。しかし対策は有効なので、全てのプラットフォーム事業者が取り組んでいくことが大事だ。

フェイクの生成技術が民主化したと先に述べた。実際に、ディープフェイク画像を作るウェブサイト[14]で「shizuoka　disaster　color」などと入力すると簡単に出てきた写真のような2枚の画像は、ぱっと見では真偽が分からない。たった3ワードを入れるだけでこういう写真が作れてしまう時代。こうした画像がこれからどんどん作られていくし、恐らく1、2年もしないうちに人間が判断つかないようなディープフェイク画像が山のように出てくるだろう。

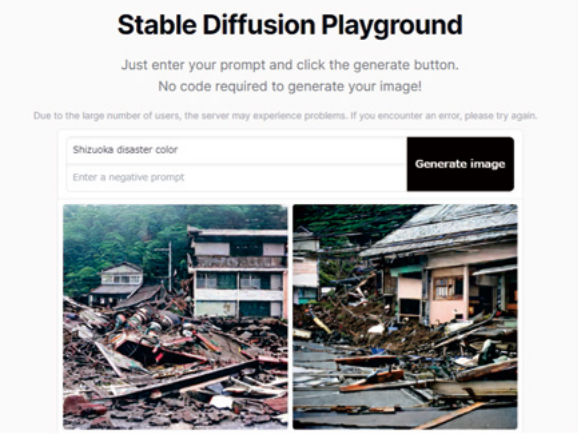


図2–2–3　　単語を三つほど入力すると簡単に出てきたディープフェイク画像

Stable Diffusion　ウェブサイト

その結果、社会が混乱するだけでなく裁判の証拠の捏造なども起こるだろう。ディープフェイク技術が民主化しているからこそ、ディープフェイクを検証する技術も民主化しなければならない。つまり、ウェブサイトに行くとこの写真はAIが作ったものかどうかを判断できるようなサービスなどが、早期に実装されることが望ましい。

メディア情報リテラシー教育も推進すべきだ。特に米国では、メディア情報リテラシー教育を推進する機運が高まっている。EUも同様だ。日本でもぜひ実践してほしい。総務省と一緒に講座の資料を作成した[15]。このように分かりやすい資料を作ってどんどん教育を推進していくことが大切だ。

13　山口真一. (2019). フェイクニュースの正体と情報社会の未来. ダイヤモンドハーバードビジネスレビュー, 2019(1). 64-73.

14　Stable Diffusion Playground https://stablediffusionweb.com/#demo

15　総務省「【啓発教育教材】インターネットとの向き合い方～ニセ・誤情報に騙されないために～」https://www.soumu.go.jp/use_the_internet_wisely/special/nisegojouhou/

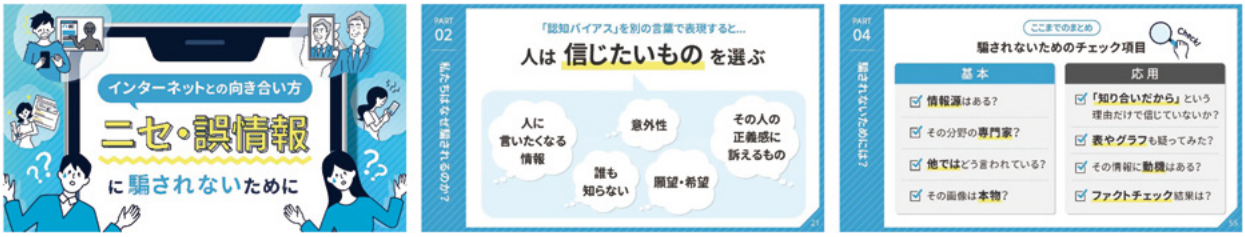


図 2–2–4　　メディア情報リテラシー教育の推進

総務省の啓発教育教材「インターネットとの向き合い方～ニセ・誤情報に騙されないために～」

アテンションエコノミーへの対策の動向にも注目したい。偽・誤情報がこれほど大量に拡散している背景には、彼らが広告でもうける仕組みがあると指摘されている。EUは、偽・誤情報対策に関連する行動規範をまとめ、関連して広告収入の停止を強く打ち出している[16]。いかにして偽・誤情報作成者が広告収入を得られないようにするかが重要で、日本でも取り組みを推進していく必要がある。

2.2.7 対策としてのファクトチェックの効果

ファクトチェックは効果的な対処法だ。2021年に菅元首相がワクチンを打ったとの報道があった際、「打ったワクチンは偽物」という偽・誤情報が出た。投稿されたときは95.0%のツイートがそれを肯定する内容だったが、その1カ月後、ファクトチェック記事が配信された後は、99.8%がファクトチェック結果を広めようとする投稿だった。つまりファクトチェックは、情報発信における主体を変更する力を持っていて、ツイッター上の言論も塗り替える効果がある。

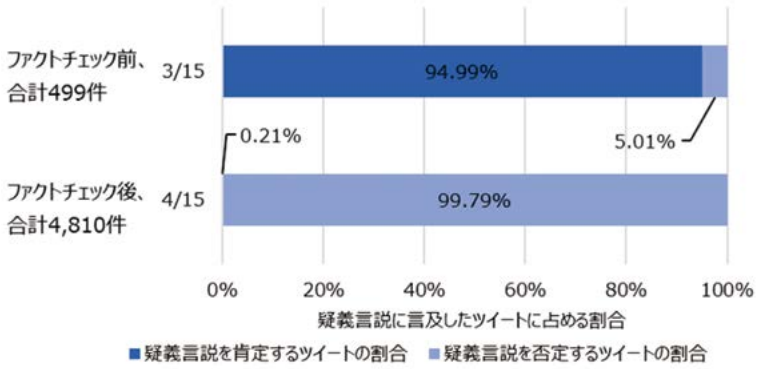


図 2–2–5　　疑義言説「菅首相が打ったワクチンは偽物」に関するツイート数

山口真一（2022）『ソーシャルメディア解体全書』、勁草書房

立場のある人が発表すると効果があることも分かっている。代表的なデマとして「ワクチンを打つと不妊になる」というものがあったが、河野太郎元大臣がテレビ出演してデマを否定した瞬間に、その話題がツイッター上で盛り上がった。彼がテレビに出演して否定する前は、実は「ワクチン、不妊」を含むツイートの60%以

16　讀賣新聞オンライン 2022年6月17日「偽情報対策のＥＵ新規範、グーグルやメタが署名…広告収入停止やディープフェイク監視求める」https://www.yomiuri.co.jp/economy/20220617-OYT1T50067/

上が不妊を信じていた。ところが、ファクトチェック後、大臣が発表した後はその割合が30%にも下がった。非常に話題になった上に、デマを信じている人が減ったことが分かる。

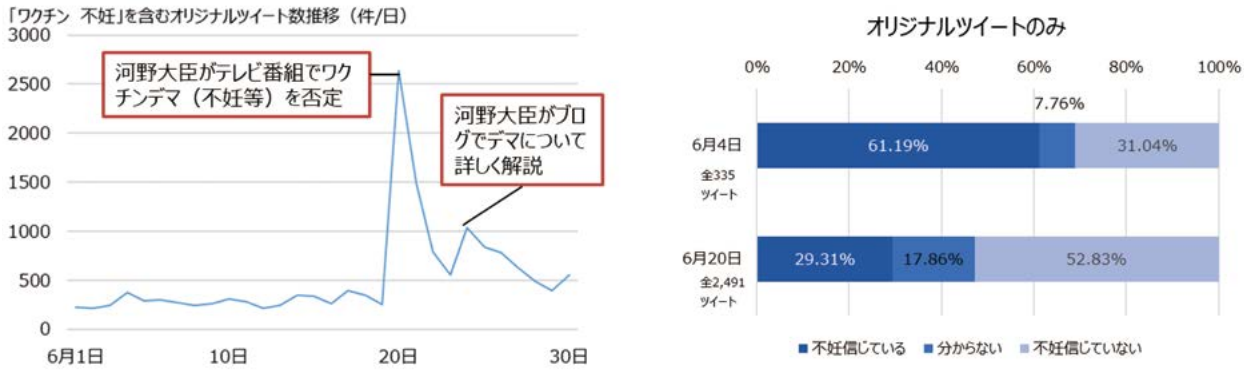


図2–2–6　　国の責任ある人の発信はデマを信じる人の減少に効果

左：「ワクチン　不妊」を含むオリジナルツイート数推移、右：ツイートの中で不妊を信じていないものの割合

ファクトチェックでもう一つ注目すべきは、マスメディアの存在だ。「情報の真偽を確かめられると便利な媒体は何だと思いますか」と問うた結果、全ての年代において人々はマスメディアにファクトチェックを期待していた（表2–2–1）。

表2–2–1　　コロナワクチン関連の情報の真偽を確かめられると便利な媒体（年代別）

真偽を確かめられると便利な媒体	20代	30代	40代	50代	60代	全年代
SNSでのインフルエンサーの投稿	9.0%	5.8%	4.4%	2.5%	1.9%	4.4%
SNSから読めるネットニュース	22.4%	20.3%	13.1%	9.2%	6.7%	13.6%
メッセージアプリのお知らせ・通知	8.4%	6.5%	4.5%	5.3%	4.2%	5.6%
メッセージアプリから読めるネットニュース	9.9%	8.3%	7.9%	7.4%	7.0%	8.0%
SNSとメッセージアプリから読めるもの以外のネットニュース	12.1%	11.8%	15.4%	16.6%	13.8%	14.2%
動画共有サービスでのインフルエンサーの投稿	6.6%	5.3%	3.0%	1.9%	3.0%	3.7%
官公庁・自治体のウェブサイト	27.3%	27.4%	33.6%	33.2%	39.0%	32.6%
官公庁・自治体のSNSアカウント・メッセージアプリアカウント	15.4%	12.5%	12.9%	11.7%	13.3%	13.0%
テレビ・新聞などのマスメディアでの報道	34.5%	30.8%	37.4%	48.9%	56.0%	42.3%
書籍	5.5%	4.6%	6.3%	4.4%	6.5%	5.5%
その他	2.8%	3.2%	4.0%	3.9%	4.4%	3.7%
特に何かで確かめたいと思わない	32.4%	36.9%	33.6%	28.7%	21.0%	30.2%

n=5569

実際に世界ではBBC、CNN、ワシントンポストなどさまざまなマスメディアがファクトチェックをしている。NHKも、「流布された誤りを正して正確な情報を不特定多数に人々に提供し直すのは、マスメディアが最も

適任である」[17]との指摘に応答している[18]。

ところが、日本のマスメディアはまだ少しファクトチェックに後ろ向きな点もある。丁寧な取材をして発信する情報は正確性が高いが、ネット上の真偽不明情報について、ファクトチェックの流れに沿ってのチェックや発表はあまりやらない。国際ファクトチェックネットワーク（IFCN）[19]という国際団体があり米国8団体、インド15団体など合わせて102団体が加盟している中、日本で加盟している団体はゼロ。韓国、台湾、香港にも後れをとっている。これについて法政大学の藤代裕之教授は「フェイクニュース対策が国際的な動きから取り残されたガラパゴス化する恐れがある」と指摘している[20]。

マスメディアが及び腰である理由は主に二つある。一つは、ファクトチェックに人員を割いていられないこと。もう一つは、検証記事は非常に時間とお金がかかること[21]。ファクトチェック団体も立ち上がっては消えを繰り返してきたが、それも人とお金が問題だった。しかし、例えば韓国では大学がハブになってプラットフォーム事業者とメディア企業が連携し、ファクトチェックを推進するプラットフォームを構築している[22]。英国BBCはテック企業と組んで偽・誤情報に対抗していくと述べている。

このように在来プラットフォーム事業者がコストを負担したり、ファクトチェック組織と大学、メディア、IT企業が連携したりと盛んに取り組まれている。日本でも大きいファクトチェック団体が立ち上がってきたし、大学、マスメディア、IT企業もある。連携して取り組んでいくことが非常に重要だ。

フランス通信社（AFP）はグーグルと組んで、ファクトチェッカーを養成する講座を作成した[23]。英語だが、申請すれば受講できる。これによってファクトチェックができるようになった人たちがファクトチェック団体で活動できるという趣旨だ。こういう連携もあり得る。

偽・誤情報は、以前からあった問題がネットが普及して大きくなったもの。社会に根付いた問題だから、撲滅はできないし、特効薬もない。しかし改善はできる。デジタル社会においてトラストを担保するには、政府、プラットフォーム事業者、マスメディア、団体など、全てのステークホルダーが連携して一歩ずつ前に進んでいくことが何よりも重要だ。

17 石川徳幸．(2018)．デジタル時代の新聞産業とジャーナリズム．情報の科学と技術，68(9)，434-439. https://www.jstage.jst.go.jp/article/jkg/68/9/68_434/_pdf/-char/ja

18 NHK 放送文化研究所 文研ブログ 2023年2月6日 #449 シリーズ「深刻化するネット上の誹謗中傷・いま何が必要なのか」(3) ～いまマスメディアに期待される役割とは https://www.nhk.or.jp/bunken-blog/2023/02/06/

19 IFCN（International Fact-Checking Network）https://www.poynter.org/ifcn/

20 藤代裕之 2019年12月28日「日本のファクトチェックに足りない3つの視点「フェイクニュース対策」ガラパゴス化の恐れ」 https://news.yahoo.co.jp/byline/fujisiro/20191228-00156700

21 藤代裕之 (2019)．フェイクニュース検証記事の制作過程～2018年沖縄県知事選挙における沖縄タイムスを事例として～．社会情報学．8(2)，143-157. https://doi.org/10.14836/ssi.8.2_143

22 総務省 プラットフォームサービスに関する研究会（第27回）2021年5月13日 資料3－1「諸外国におけるファクトチェックの取組について」(みずほリサーチ&テクノロジーズ）https://www.soumu.go.jp/main_content/000749421.pdf

23 AFP Digital Courses “Digital Investigation Techniques” https://digitalcourses.afp.com/

2.3 ［振る舞い予想・対応可能性］システムのトラスト

中島 震

システムの振る舞いの予想、それに対応することについて話題を提供する。ここで、システムとはどういうものを考えているかというと、例えばインターネットサービスやAI、機械学習のソフトウェアのことである。振る舞いの予想を行う上で、ソフトウェアシステムの開発に着目する。二つの観点があり、内容のベースは『ソフトウェア工学』と『AIリスクマネジメント』である。

2.3.1 ソフトウェア工学における開発と利用

ソフトウェア工学はソフトウェアシステム開発を対象とする。図2–3–1に示すように、発注者あるいは利用者が暗黙に持っているような要求を開発者は要求仕様としてまとめ、それを最終成果物であるプログラムとして作っていく。ここでは、構築の過程と、最終成果物がそもそもの要求仕様を満たしているかどうかを検査する過程がある。

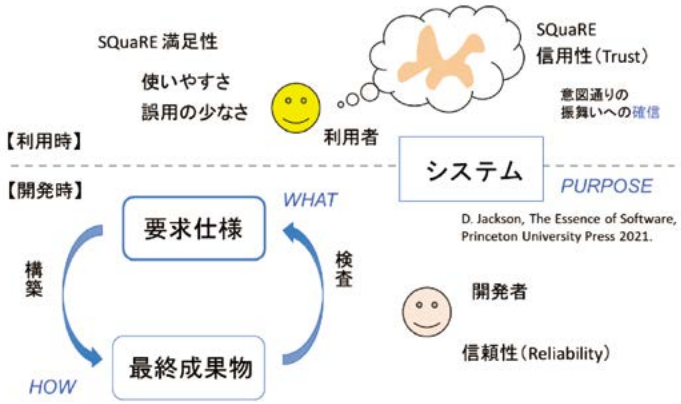


図2–3–1　　ソフトウェア工学における開発と利用

一般的に要求仕様を考えるときには、どのように作るかということはなるべく考えないで、純粋に利用者あるいは発注者が考えている要求をなるべく素直に書き表したいということから、WHATを表すというふうに呼ぶ。一方、プログラムを作るときには細かなデータ構造やアルゴリズム、あるいは基盤ソフトウェアつまりOSなどの使い方を熟知して作らなければいけないので、HOWの部分が重要になる。そして、構築の際にはWHATに準じた形で物を作る、あるいは検査のときには作ったものがWHATを満たしているかを調べることになる。したがって、まずはWHATのところが重要になり、システムに期待する振る舞いをきちんと予想し、対応することが重要になってくる。

開発が終わると、利用のフェーズになる。利用者から見ると、システムが期待する機能、振る舞いを示すかが重要になる。つまり、開発時点での要求仕様を正しく反映しているか、うまく作っているかという話になる。そして、利用者から見るとシステムに対する満足性としては、使いやすいとか、誤用が少ないというような性質として見えてくる。したがって、利用者の満足性を高めるには、利用者と一緒に物を作る、あるいは利用者がシステムに対してどのようなイメージを持つかを知る、いわゆるユーザーエクスペリエンスというものが重要になる。

そういうものをひとまとめにして、SQuaREというソフトウェアおよびシステムの品質のモデルがあり、そこ

では満足性という言葉を使っている。そのSQuaREの満足性の副品質特性に信用性というものがある。システムが意図どおりの振る舞いをどの程度示すかということに対する利用者の確信度のことを信用性と呼んでいる。「信用性」という言葉はSQuaREがJIS規格になったときの日本語であるが、元の英語ではTrustであった。ここでトラストというキーワードが、ソフトウェアのシステムでは、このようにユーザーから見た視点として使われていることを記憶にとどめておいてもらいたい。

そして、もう一度、システム開発の側に戻ってみると、かなり品質よく要求仕様が作られていたとしても、それが利用者の信用性に直接寄与するとは限らない。開発者が考えるシステムの目的と利用者が感じる目的の間に不一致があると、システムがよい機能を発揮し、それが信頼性が高く作られており、不具合なく動いていたとしても、利用者にとっては満足度が高くならない。その辺についてはDaniel Jacksonがその著書において、WHATとHOWだけではなくてPURPOSEを重要視するといっている。しかし、そうは言ってもPURPOSEというのも開発者が考えるべきことであり、振る舞い予測や対応に関する開発者の能力の表れになる。

2.3.2 要求工学における要求の取り出しと形式化

要求工学というのは、発注者がイメージしているものをいかに要求仕様として具体化するかということである。要求仕様としてうまく整理できれば、その後はその仕様に従ってシステムを構築するという通常の開発の活動になる。要求工学の場合には要求仕様を得る過程（Requirement Acquisition）を二つに分けて考える。一つはElicitationで、どのようにして要求を引き出すかという部分である。もう一つが引き出した結果を開発グループで共有するための要求の形式化（Formalization）である。

最初のほうのElicitationには二つの問題がある。要求を持っている人が本当に自分の要求を把握しているかということが一つ目である。通常は要求を出発点にElicitationのことを考えるが、そうは言ってもElicitationのときに一番厄介なのはKnown Unknownの壁である。つまり、何が分かっていないかということが分かっていたとしても、分かっていないところを何らかの形で埋めないと形式化ができなくてシステムの開発には至らない。ところが、Known Unknownというのは何を言っているかというと、知ることの限界を言っているわけである、つまり全てのことを知ることはできないということであり、Known Unknownのままではシステムの開発はできないので、最悪の場合、システムの開発者が自分の都合のよい形で形式化してしまう可能性がある。これは要求工学において常に問題になっていたことである。このように何が問題かということは割と整理できているが、完璧な解決策はまだないという状況で、Known Unknownのことを考えると永遠に解決策がないということになる。

2.3.3 システムが及ぼす社会的、倫理的な影響

今までのシステムを作るときに想定するのは直接の利用者である。直接の利用者について考えることで、振る舞いを予想し、対応していた。しかし、AIなどのシステムでは影響をかぶる当事者が直接の利用者以外の人であることがある。そのときの影響は、社会的な不利益や倫理的な不利益というものが考えられる。簡単にいうと、利用者と違った第三者がいて、その両者の間では利害が同じとは限らない。したがって、利用者だけの要求やUXだけを考えていたこれまでの要求工学とはまたもう少し違った観点の要素を入れる必要がある。

さて、社会的、倫理的な影響というものを一言で言うならば、自動化、AIの技術によって何らかのバイアス、つまり偏見のようなものが増幅されることが問題になっている。例えば、数学教師の公募があり、応募書類で人を評価しようとする、そんなプロファイリング技術を使うシステムがあったとする。そのとき、直接の利用者から見ると多数の応募者の中から、何らかの観点でよい人を選べばよいということになる。ところが、よくある話としては、学習データには偏りがあり、特に伝統的に数学教師には男性が多いという傾向が仮にあったと

すると、このシステムの評価結果が女性の応募者に対して不利益をもたらすかもしれない。この場合は、性別という点で、評価の公平さを欠くことになる。仮に、学習過程でのデータの取り扱いに問題がなく、学習システムがきちんと動いたとしても、学習に使ったデータは私たちの社会を反映しているので、そういったステレオタイプが問題になり、倫理的な不利益が起きる可能性がある。このような事例は数多く報告されている。

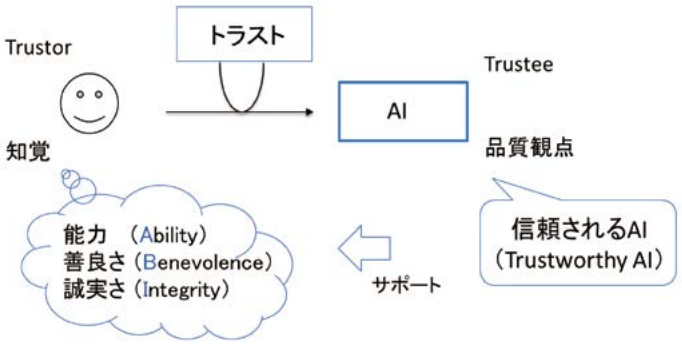
2.3.4 AIとトラスト

未知の複雑な技術というのは、私たちの側に経験がないので、不確かさが大きい。つまり、非常に大きな便益を得られるかもしれないが、危害が発生するかもしれないという不安がある。

そういった不安を解消するための一つの考え方としてトラストというものがある。新技術に対して、トラストすることによってリスクをあえて取る、つまり被害があるかもしれないが、便益のほうが大きいだろうと考えてあえて新技術を使っていくという考えである。ここではトラストを社会受容性から考える。

このトラストは、先ほど出てきたSQuaREが規定しているトラストとは若干意味が違うと思われる。SQuaREは、どちらかというと複雑さの軽減ということを考えていて、大規模なシステムや、あるいはメカニズムが分からないようなシステムが、自分が想定するような振る舞いをしてくれるという確信の度合いのことをトラストと呼んでいた。一方、ここでのトラストというのは、リスクをあえて取る、不確かさな状況をどのようにして自分の頭の中で納得するか、ということである。

そういう観点から、図2–3–2に示すABIモデルというものが提案されている。1995年に出された論文がベースになっており、能力（Ability）、善良さ（Benevolence）、誠実さ（Integrity）というの三つの性質を持つことを人が知覚すると、対象との間にトラスト関係が生じると考える。能力というのは、期待どおりの振る舞いをして機能を果たすことができるということである。善良さというのは自分に悪さをしては困るということであり、先ほどの影響をかぶる第三者という話があったが、あのような形で倫理的なあるいは社会的な不具合というものをもたらしては困るということである。そして、誠実さというのは、システムあるいは対象は決められた規則どおりにフェアプレーで（公正に）動いてくれなければいけないということである。



Roger C. Mayer, James H. Davis, and F. David Schoorman, An Integrative Model for Organizational Trust, The Academy of Management Review 20(3) , pp.708-734, 1995.

図2–3–2　　トラストのABIモデル

それで今、Trusteeの部分をAIと考えたときに、人がABIを知覚するためには、どのような品質を満たせばよいのかということが次の議論になる。それをTrustworthy AIと呼ぶ。

さて、ここでTrustworthy AIの中身を見ていく。図2–3–3に示すように、複数のステークホルダーがいて、システムやサービスだけではなくて、ステークホルダーも含めた集団との間にトラスト関係が成り立つというよ

うに考える。

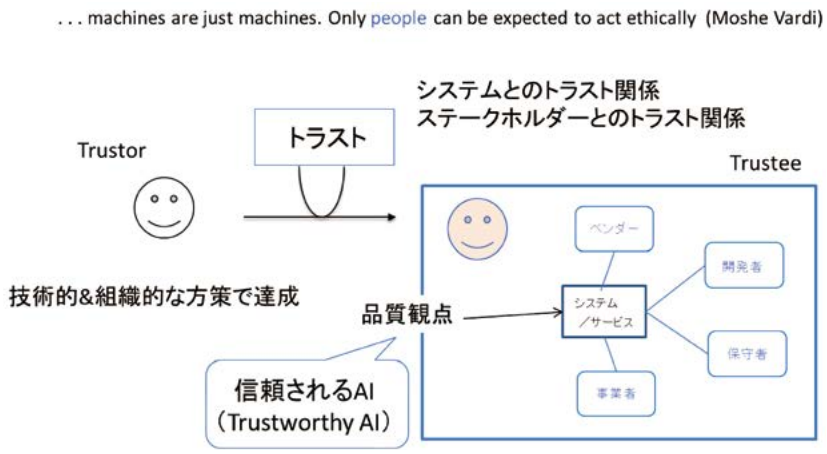


図2–3–3　　トラストと信頼されるAI

そこで、Trustworthy AIが持つべき品質を考えると、それは技術的な方策で達成することもできるだろうし、あるいはベンダーとか事業者、保守者、運用者などの組織的な方策で達成するものもあるかもしれない。そこはあまり区別せずに、とにかくTrustworthy AIが持つべきである、期待する品質の観点にはどんなものがあるかということを次に考える。ただし、Trustworthy AIもソフトウェアである。

ここでソフトウェアシステムの品質について歴史的に概観する。1980年代から90年代にかけては、信頼性が最も重要であった。装置の故障に対するフォールトトレランスが重要な課題だった。次に、システムの機能が複雑になると、アプリケーションのレベルでの品質、例えば安全性が重視されるようになった。要求仕様を守った上で、外界に対して悪影響を及ぼさないということが安全性の考え方である。次にインターネットが導入され、システムがネットワーク化されると、CPSが持たなければいけない品質特性として信頼性、安全性に加えて、回復性やサイバーセキュリティー、プライバシーを2016年にNISTが提唱した。一方、2002年にはマイクロソフトのBill GatesがTrustworthy Computingとして、Business Integrityに言及し、誠実なビジネスを行うことが顧客のトラストを高めると論じた。先のABIモデルに似たニュアンスである。

2.3.5 AIと品質

ソフトウェアをシステムとして見たときに、品質の観点がいくつかあるが、信頼されるAIを考えたときには、ソフトウェアとしての品質観点、従来のCPSを含むような高度なソフトウェアシステムが持つような品質観点、そして、最後に社会的、倫理的な品質観点、というよう分類できる。 AIの場合は、図2–3–4に示すように、いろいろな観点が複雑に入ってくると考えている。

この表は私がまとめたものであるが、NISTから出ているAIリスクマネジメント・フレームワークで掲げられている品質観点と見比べたところ、大体同じようなことを見ている。このような観点が今後の重要な品質観点であると言って良いだろう。

分類	品質観点	T	S
バニラ AI (Vanilla AI)	モデル正確性 (Model Accuracy)	△	
	モデルロバスト性 (Model Robustness)	△	
説明可能 AI (Explicable AI)	透明性 (Transparency)	△	○
	解釈可能性 (Interpretability)	△	○
ロバスト AI (Technically Robust AI)	信頼性 (Reliability)	△	
	安全性 (Safety)	▽	
	サイバーセキュリティ (Cybersecurity)	▽	
	回復性 (Resilience)	▽	
倫理的な AI (Ethical AI)	公平性 (Fairness)	▽	○
	プライバシー (Privacy)	▽	○
合法的な AI (Lawful AI)	アカウンタビリティ (Accountability)		○
	適合性 (Conformity)		○

T：製品品質（△）・利用時の品質（▽），S：組織上の方策

図 2–3–4　　信頼される AI の品質観点

（「中島震：AI リスク・マネジメント、丸善出版 2022」表 4.1 より）

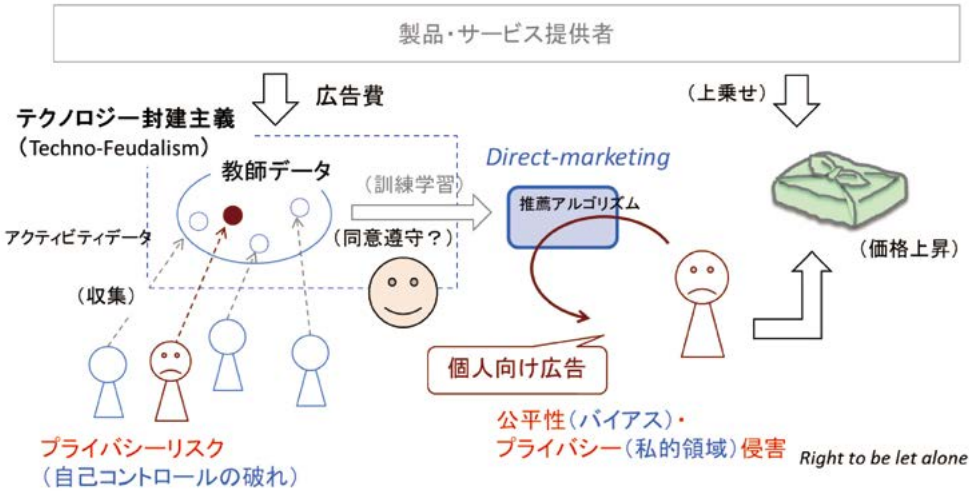


図 2–3–5　　アルゴリズム不正義

次に、Ethical AI というものについてもう少し見ていきたい。簡単な例として、図 2–3–5 のような、個人向け広告を出すということを考えてみる。実はここには、いろいろと難しい問題がある。確かにこの人にとっては非常によい情報、簡潔な情報が効率よく得られるわけであるが、二つの問題があると言われている。

一つは公平性の問題である。特定の顧客のさまざまなセンシティブな特性を使って個別の広告を出すのであるから、センシティブな特性がこの人に対して本当に善良に解釈されているのかどうかということは少し考えてみる必要があり、公平性の問題、バイアスの問題がある。もう一つは、個人にカスタマイズされているということは、実は私的領域を侵害していることになる。Right to be left alone ということをあとで述べるが、これは基本的な人権なので、それに対してどのように考えるかということである。

もう一つは自己コントロールの問題である。非常に精度の高い、よい推薦アルゴリズムをつくるためにはさまざまなデータを集めているが、そのデータは本人とは別の人から集めてくることになる。すると、データを集められる人の観点から見ると自分自身のアクティビティーデータを提供していることになるが、その提供したデータというものがどこで誰がどのように使っているのかと言うことが分からない。つまり、自己コントロールの破れが生じてしまうことになる。

この辺のことはある種、同意の話になるが、同意の考え方というのは、世界でもいろいろな考え方に分かれ

ている。韓国では、こういった個人情報を無断で使っているということから、プラットフォーム企業に対して、100億円規模の課徴金を課したというニュースもあったと記憶している。プライバシーの問題、個人の領域の侵害、つまり私的領域の侵害という意味でのプライバシーの問題と自己コントロールの破れというプライバシーリスクの問題点がある。

公平性とプライバシーは、倫理的なAIとして非常に重要な側面であるが、公平性の話よりはプライバシーのほうが歴史的に確立した概念になっている。それを見ることによって法規制とAIの関係が少し理解できると思われる。

プライバシー権は19世紀に生まれた概念であるといわれ、例えば芦部信喜の著書「憲法」によると、日本国憲法第13条の幸福追求権が日本ではプライバシー権の基になっている。これを分解すると、一つはRight to be left alone、一人で放っておいてもらう権利であり、もう一つは情報プライバシー権と呼ばれる、自己に関する情報をコントロールする権利になる。最近は自己コントロール権のほうに注目が集まっている。

自己コントロール権の法律の例としては、欧州のGDPRがある。さらに、欧州ではAI-ACTが議論されていて、GDPR流の見方と私的領域を侵害する問題の両方を考えている。また、少しプライバシーの考えとは離れるが、欧州ではデータ主体の話だけではなく、企業が持つデータやその権利に関してデータアクト（Data Act）が議論されている。AIなどのデータ利活用の高度なソフトウェアが今後考えていかなければいけない関連法案、法規制の一つになると思われる。

技術の流れとしてはAIにおいても、公平性やプライバシーに関わる技術が研究がされている。しかし、これが必ずしも両立しないという報告がある。公平性を高めるとプライバシーが低下する、あるいはプライバシーを高めようとすると公平性が低下したり、ばらつきが増えたりするという関係にある。したがって、こういった技術的な問題点を考えた上で、倫理的なAIとして何を重視するのかというものは一段上のメタなポリシーのところで考えて、要求仕様に取り入れることを考えなければいけない。

2.3.6 おわりに

振る舞い予想・対応可能性を向上させることで、AI技術に対するトラストを形成しようということになる。ソフトウェア工学の方から考えると、WHATの問題の部分とHOWへの影響の部分というものに分けて考える。WHATの問題については、Known Unknownの壁というものが厳然として存在するし、社会的な影響も大きくなっているので、それらへの対応を考えることが非常に重要になっている。 HOWのほうというのは、突き詰めていくと結局のところ、どういう学習データを使うかという話になる。バイアスへの対応を考え、期待するような形でデータを整備し、システムが作り込まれているかを確認する方法論としては、『AIQM品質マネジメントガイドライン』に記載されているチェックの観点を意識することによって、改善されると期待している。

結局のところ、運用中にトラストが低下する、つまり、振る舞い予想が外れるとか、対応ができなくなる、というような状況が生じることは十分に考えられる。そのときに、トラストの低下を運用中に検知し、俊敏に対応する技術上ならびに組織上の仕組みが期待されている。

3 ｜ トラスト３側面の融合・分野横断の必要性

3.1 トラスト３側面の捉え方

大屋 雄裕

同じようにトラストといっても人によって意味していることが違う。これが、このテーマで議論を始めたときの最初の発見だった。それを三つにまとめたのが「トラスト３側面」（対象真正性、内容真実性、振る舞い予想・対応可能性）だが、今度はそれらがどのように関係しているのかを位置付けてみたい。

トラストの議論をするときの目的は何かというと、われわれが安心して社会生活を送ることだろう。「安心して」というところがポイントで、ゼロリスクを追求するというルートもあり得るが、コストがかかり過ぎて非現実的である。例えば環境規制では、ある時期「予防原則」として、予見されていないリスクも含めて可能な限りリスクをゼロにすることが提唱されたが、10年ぐらいたつとやっぱりむちゃだったと反省されるようになり、コストパフォーマンスが強調されるようになった。

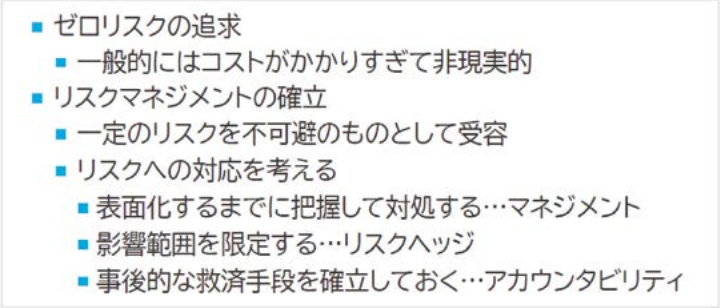


図3-1-1　　安心して社会生活を送るためには？

コストパフォーマンスを考えることは、リスクマネジメントをするということだろう。一定のリスクは不可避なものとして受容する一方、リスクが露呈したときのために対応を考えなければいけない。

その方法はいくつか考えられる。一つはリスクが表面化するまでにそれを把握して対処する狭義のマネジメント、ずっと注視する方法である。次に、リスクが露呈したとしてもそれを一定の影響範囲までにとどめるような限定策を講じる、リスクヘッジをしておくことも一つの方策である。また、何らかの形で事後的に救済手段が得られるのであれば、それでもいいという割り切り方もある（アカウンタビリティー）。

株式投資を例に考えてみる。マネジメントは、リアルタイムで株価を注視していて、値下がりが一定の許容水準を超えたら売るということになるだろう。このときに必要なのは透明性である。次に、リスクヘッジは、分散投資やロング・ショート戦略など、リスクが巨額になり過ぎない形で影響範囲を限定することによってリスクに対応していくことになる。三つ目のアカウンタビリティーについては、証券会社が破綻して預けておいた金がなくなることを防ぐために、現実の法制度としては顧客から預かったお金は証券会社の独自の資金とは別に分別管理するように定められていたり、それでもカバーし切れない損失が生じたときのために投資者保護基金がつくられて、証券会社が一定のお金をデポジットすることになっている。「きかんしゃトーマス」の「じこはおこるさ」という愉快な歌のように、事故が起こることを前提として救済制度をつくるという対策だろう。

ただ、このようなリスクマネジメントのためには、リスクに関する情報が間違っていないことが不可欠である。異なった対象ではなく、まさにその対象を扱っていることが保証されている必要がある（対象真正性）。それ

から、得られた情報が捏造されたものでない、あるいは虚偽ではないという意味での内容真実性も必要になる。事実と異なる情報ではないことがどこかで証明されている必要がある。このように考えてみると、対象真正性や内容真実性は、三つ目の振る舞い予想・対応可能性の前提になるという位置付けができるだろう。

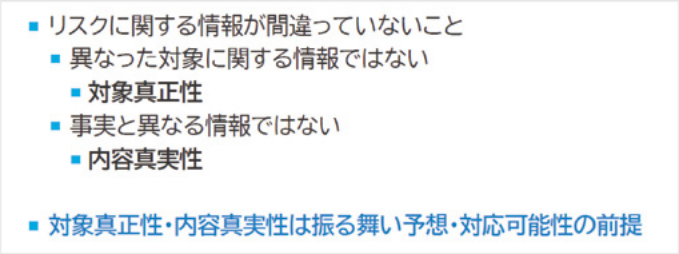


図3–1–2　　前提となるもの

リスクマネジメントという観点からは、振る舞い予想・対応可能性は多様であり得ることを強調したほうがいいだろう。例えば、出力を上げると暴れる芝刈り機を考えよう。そのようなリスクが対象真正性や内容真実性から明らかになったとして、われわれがそれに対応する施策は何通りもあり得る。出力を限定して運用する、すなわちリスクが顕在化しないように回避することも一つの方策だろう。また、暴れたとしても腕力で押さえつければいいというジャイアニズム（ジャイアン主義）のような考え方もあり得る。リスクを許容するという方策である。それから、この例ではお勧めしないが、芝刈り機が暴れてどこかにぶつかって損害が出ることを前提として保険に入るというような形で、リスクに対する対応を考慮しておくという施策も考えられる。

このようにさまざまな方策があり得る以上、振る舞い予想・対応可能性の内容は客観的に決まらない。あえて言うと、対象真正性とか内容真実性はある程度、物理的・客観的に決まるし、それを超えたところでも社会的な合意形成機能があって調整ができるだろう。ニュースのファクトチェックなどは、本当の客観性というよりは、今述べた社会的合意形成システムが存在している例。そういうものも及びにくいようなものとして振る舞い予想・対応可能性があるのではないか。

さらに言うと、これは人によって違う。戯れ言ではあるが、犬好きな人はリスクが存在しないことを好み、相手に言うことを聞かせたり、相手がこちらの思った対応を取ることを尊重するタイプだろう。それに対して、猫好きは相手が自分の予想外のことをやるから楽しいと感じている、リスクを楽しむタイプだろう。本当にそうかはともかく、そういうタイプの違いも当然あるだろう。

この辺りの多様性を織り込んで研究しないと、トラストは分からない。その意味では、対象真正性や内容真実性に関する、語弊はあるが、理系的なアプローチとともに、社会科学的・人文学的な手法が重要になってくる側面があると考えることができる。

あと、振る舞い予想・対応可能性の内容によって、想定される対象真正性とか内容真実性の水準も変わってくるという点にも注意したほうがよい。例えばリスクを回避しようとする人にとっては、その閾値がどこかというのが重要である。芝刈り機の出力が4段目で暴れるのか、5段目で暴れるかは重要な差として効いてくる。しかし、保険でカバーするというタイプの人にとっては大体の事故率が分かっていればよく、閾値は関係ない。あるいはお風呂のお湯の温度を考えると、40℃か50℃かはクリティカルに違うが、20℃か30℃かの違いはどちらにしろぬるくて入れないので関係ない。そう考えると、振る舞い予想・対応可能性のところにかなりの多様性とかキャラクターが出るため、それを取り込んだ形で研究する総合的な取り組みが必要になる。

トラスト3側面の相互関係を表にまとめた（図3–1–3）。どちらかというと対象真正性や内容真実性がリスク判断の前提であり、客観的・外在的な検討が可能であるのに対して、振る舞い予想・対応可能性はリスク判断の内容をなしており、主観的・内在的性格を持っていると言えるのではないか。それゆえに、それらを取

り込んだ総合的な研究をしないといけない。

対象真正性 内容真実性	リスク判断の前提	客観的・外在的
振る舞い予想・ 対応可能性	リスク判断の内容	主観的・内在的

図3–1–3　　トラスト3側面の相互関係

【質疑・討議】

福島：内容真実性には結構主観的な側面があると思っていたが、どのように理解すればよいか。

大屋：あくまで相対的な違いにすぎない。内容真実性も例えば今日、東京都に雨が降ったかどうかにはあまり主観性はない。雨が降ったということを定義するところで社会的合意が成立すれば、検証条件は客観的にある。もちろん、先ほどのフェイクニュースの話では、決して客観的ではなく、起きた事実が客観的であったとしても、その理解の方法は多様である。だから、もちろん、そこにかなりの主観性は出てくる。ただ、そのような交ざり方は、恐らく対象真正性でもある。中川裕志先生の問題提起で出てくるが、人がアバターを使ってコミュニケーションするようになったときに、アバターが本体なのか、人間が本体なのかは、事実として決まるというよりは社会合意によって決まるだろう。あるいは人体のエンハンスメント技術がある程度、実用化され始めているが、そうやって人間の肉体が外側に拡張されたときに、どこまでを人と考えるか、どこまでが対象として検証の対象になるのかは、社会的に合意として決まってくることだろう。どちらも両側面は交じり合うが、あえて強調すれば、こういう形でかなりの差があるという話として考えてほしい。

3.2 人文・社会系での取り組みと分野横断の必要性

小山 虎

3.2.1 トラスト研究の多様性

人文・社会系のトラスト研究の系譜について紹介する。2018 年に出版した『信頼を考える－リヴァイアサンから人工知能まで』（小山虎編著、勁草書房）では、哲学の社会思想に始まり、行動科学、政治学、社会心理学、ビジネス、教育、医療、ロボット、障害者福祉、ヘイトスピーチ、ヒューマンエージェントインタラクション（HAI）における信頼を取り上げ、信頼研究の全体像を大まかに示すことを試みた。

図 3–2–1　『信頼を考える』（同書から表紙と目次を引用）

このようにさまざまな研究分野で信頼が扱われている。共通点もあれば、違う点もある。人文・社会系に限っても、用いられている信頼の定義には幅がある。これは信頼という現象そのものの問題なのか、それとも、われわれが使う言葉の問題なのか、いろいろと議論の切り口がある。

分野	定義
経済学	「［信頼は］、1人ないし複数の行為者が特定の行為を遂行するという一定のレベルの**主観的確率**であり、彼らの行為をチェックすることができる以前に（あるいはチェックすることができる能力とは独立に）、その行為が自分自身の行為に影響を与える状況で形成される」(Gambetta, D (1988). Can We trust trust? In Trust: Making and Breaking Cooperative Relations: 217)
経営学	「［信頼は］、他者の意図や行動についてのポジティブな**期待に基づきリスクを受け入れる**意図を含む心理状態である」(Rousseau, D., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. The Academy of Management Review, 23(3): p. 395)
社会学	「［信頼は］欠けている情報を内的に保証された安全性に置き換えるのであり、［自分に］利用可能な情報を超えて、行動の**期待を一般化**することにより、**社会の複雑さを減少**させる」(Luhmann, N. (1979) Trust and Power: p. 93)
	「［信頼は］、他者の**誠実さや愛**あるいは抽象的な原理への信念を表すような、人やシステムが一群の結果や出来事を実際にもたらすという**確信**」(Giddens, A. (1990) The Consequences of Modernity: 34)
動物行動学	「信頼は、他者の**誠実さ**または協力への依存、あるいは少なくとも他者があなたを欺かないという**期待**、と定義される」(de Waal, F. (2009). The Age of Empathy: Nature's Lessons for a kinder Society: 167)
哲学	AがBはCすると信頼するのは、(1) AはBがCすると**期待**し、(2) このAの期待(1)が、Bが自分の関心を叶えようという動機に基づいているというAの信念か知識に基づいているとき (Hardin, R. (1991) Trust and Trustworthiness, Russell Sage Foundation)
	AがBはCすると信頼するのは、(1) AはBに重要事Cを任せ、(2) AはどのようにCを扱うのかのコントロールをある程度Bに許し、(3) AはBがCを扱うことができると**確信**しており、(4) Aは自分に対するBの**善意**に**確信**を持っているか、少なくともBの悪意や無関心を予期しないとき (Baier, A. (1986) Trust and antitrust, Ethics 96)
	AがBはCすると信頼するのは、(1) AがBの**善意**に対する**楽観的態度**を持ち、(2) AはBの予期される行動Cに対するBの能力に対する**楽観的態度**を持ち、(3) Aは自分が頼りにすることを認識することによって直接BがCするように動機付けられると信じているとき (Jones, K. (1996) Trust as an Affective Attitude, Ethics 107)
	AがBはCすると信頼するのは、(1) AはCの配慮にあたってBがある**社会的規範**に内的にコミットしていると**期待**し、(2) Aは「BがCの配慮にあたってAによって想定されている**社会的規範**を認識し、また、その規範が何を要求しているかを理解することができる」と**確信**しており、(3) AはBが自分に課せられた**規範**にしたがって行為することができると**信じている**とき(Mullin, A. (2005) Trust, Social Norms, and Motherhood, Journal of Social Philosophy 36)

図3–2–2　　さまざまな分野の信頼の定義

3.2.2 囚人のジレンマと信頼

いくつかの切り口はあり得るが、情報量の観点から、囚人のジレンマに着目して信頼研究の系譜を検討する。囚人のジレンマのような実験を用いて信頼を分析するのは、一部の分野では標準的な方法になっている。個人が合理的であるだけでは必ずしも協調行動を取るメリットがあるがどうか分からないが、相手の取り得る行動を踏まえた上で合理的に考えれば協調行動を取る方が有利な状況がある。信頼があれば、協調行動を取ることが合理的な選択になる。恐らくはそれをもって社会が形成されるだろう。反対に、信頼がなかったら人々は協調行動を取らない。それぞれが自己利益のために行動し、社会は生まれないだろう。

3.2.3 信頼研究の始まり：社会契約論

囚人のジレンマが関わる信頼研究の起源は、少なくとも、17世紀の社会契約論にまでさかのぼることができる。 Thomas Hobbes（トマス・ホッブズ）によれば、人間というのは自身の生存を目的としている。そのため、自然状態では自己保存のために争いが起こる。これが万人の万人に対する闘争といわれるものである。

社会契約論は、簡単に言うと、社会秩序はいかにして可能なのかという問題を扱う分野を指す。これはホッブズ的秩序問題と後に呼ばれるようになる。ホッブズによると、社会契約を結べば社会は作れるわけだが、自然状態で闘争している段階から、信頼ないし協調行動を取るという途中段階がある。

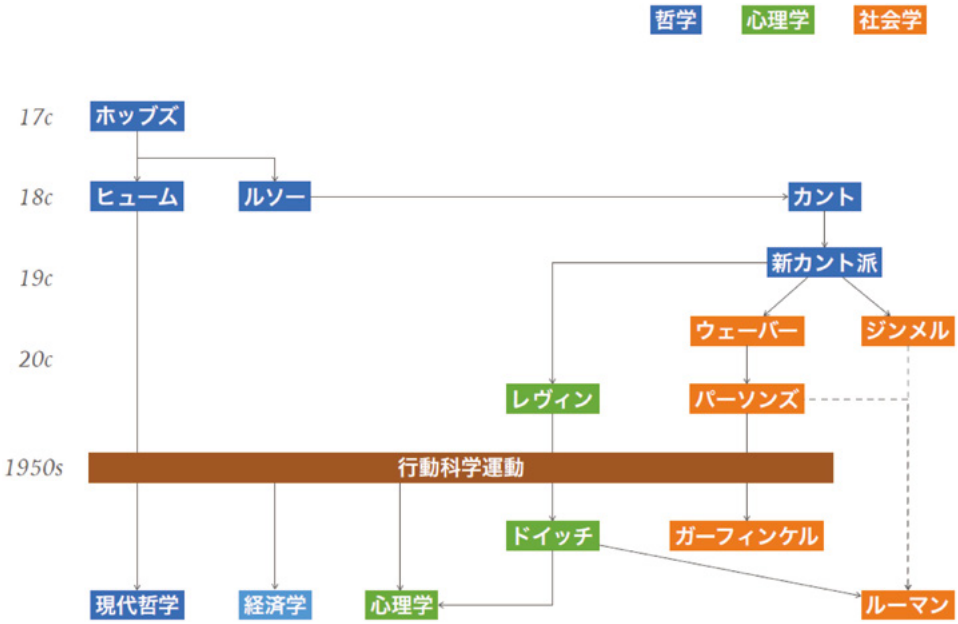


図3–2–3　信頼研究の系譜（『信頼を考える』から図を引用）

注意しなければならないのは、ホッブズが現在あるような信頼研究や囚人のジレンマに関係するような信頼を中心的に研究し始めたわけではないという点である。特に、この時代までは、神や宗教の影響も強く、信頼を信仰や信用とどのように区別するかも難しい。ホッブズは信頼（Trust）と信用（Believe）や信仰（Faith）を区別していない。

18世紀には、David Hume（デイヴィッド・ヒューム）、Jean–Jacques Rousseau（ジャン＝ジャック・ルソー）、Immanuel Kant（イマヌエル・カント）といった哲学者たちが、独自の見解を出している。ヒュームは共通利益についての一般的な感覚（コンヴェンション）が、社会秩序の形成の鍵だと考える。ルソーは、ホッブズが社会秩序はどうやって生まれるのかを考えたのに対し、社会秩序はいかに維持されるのかを考えた。ホッブズが信頼に関心を持っていたのに対し、ルソーは不信を考えたと言ってよいかもしれない。カントは制度や役割に基づく相互行為の継続によって秩序が保たれると考えた。

3.2.4 哲学から社会学・心理学へ

19世紀になると社会学が生まれる。1838年、フランスの社会学者（実証主義者）Auguste Comte（オーギュスト・コント）が「社会学」という用語を導入した。19世紀後半から20世紀にかけて、社会学者と呼ばれる人たちが登場してくる。その代表として、フランスのEmile Durkheim（エミール・デュルケーム）や、ドイツのMax Weber（マックス・ウェーバー）、Georg Simmel（ゲオルグ・ジンメル）などが挙げられる。

20世紀初頭のドイツ社会学はドイツ哲学（特に新カント派）からの影響が強かった。19世紀終わり頃に、新カント派の復興運動がドイツで起こった。その結果、19世紀の終わりから20世紀の頭のドイツの社会学というのは、カントを背景にして社会学を考える、あるいはカントの用語やカントの位置付けみたいなものが中で意識されることとなった。

信頼に関していうと、先ほど紹介したホッブズ的秩序問題が登場するのは、20世紀になってからである。ホッブズ的秩序問題と命名したのは、アメリカを代表する社会学者Talcott Parsons（タルコット・パーソンズ）である。ドイツで社会学を修めたパーソンズは、ホッブズからカントを経てドイツ社会学へと引き継がれた問題意識を受け継ぎ、価値や規範の共有が秩序を維持する要素であると考えた。

信頼の系譜を考える上では、心理学も重要である。今日、われわれが知っているような実験心理学は、1879年にドイツの心理学者Wilhelm Wundt（ヴィルヘルム・ヴント）が実験心理学研究室を開設したのが始まりとされている。ドイツの心理学者Kurt Lewin（クルト・レヴィン）はナチスに追われてアメリカに移住し、MITでグループダイナミクス研究センターを設立し、以降、社会心理学が広まった。ヴントとレヴィンも、カントを背景にして心理学を考えた。

20世紀になるとアメリカの社会心理学者Morton Deutch（モートン・ドイッチ）が登場した。彼はレヴィンの教え子であり、「紛争（解決）」や「協調」を主な研究テーマとした。彼は「囚人のジレンマ」を用いた信頼研究を最初に実施し、社会集団内の紛争解決の要因として信頼を位置付けた。その後、信頼を主観的確率（ないし期待）として理解するのが一般的になっていった。主観的確率として信頼を考えるのは、一つの標準的な考え方であり、冒頭で言及したさまざまな分野の信頼の定義にも入っているが、これはドイッチの囚人のジレンマを用いた信頼研究とともに広がっていったと考えられる。

3.2.5 行動科学運動

囚人のジレンマをベースにした信頼の考え方が広がっていった背景には、1950年代の行動科学運動があった。これは社会学や心理学が中心となって社会科学系の諸分野で学際研究を推進する運動を指す。パーソンズは行動科学運動に積極的に参加した社会学者の代表である。アメリカ国立科学財団（NSF）の設立時（1950年）に、心理学者、社会学者、人類学者が「行動科学」という名称で自分たちの分野も加えるように要請した。最終的に行動科学の名称としてではなかったが、NSFの中に社会科学系の分野が設置された。また、行動科学運動の広まったもう一つのきっかけとして、フォード財団が1950年代に行動科学プロジェクトを実施し、スタンフォード大学やハーバード大学、シカゴ大学に莫大に助成金を投入した。その結果生まれた組織の一例が、スタンフォード行動科学先端研究センターである。このように、行動科学運動は、現在の学際研究や異分野融合、分離横断と呼ばれるはしりとなったが、1960年代には廃れた。

この時期に行動科学運動が起こった背景には、第2次世界大戦中の軍産学複合体がある。戦時中は、大学が軍から委託された学際研究を行うことが当然のように行われていた。冷戦を理由に、戦後もこの体制は継続した。太平洋戦争が始まる前、アメリカでは国防研究委員会が設立されたが、戦後、この組織は解体され、後続組織として紆余曲折の結果生まれたのがNSFである。ベトナム戦争の頃まで、第2次世界大戦中に広まった軍産学が集まって学際研究を行うという体制は続いた。行動科学運動の目的は、軍（政府）からの研究委託が途切れないようにするためであった。

囚人のジレンマは、そもそも、紛争解決に関する研究として使われていた。1950年、カリフォルニアのランド研究所で実施されていたゲーム理論の研究から囚人のジレンマ・ゲームは生まれた。ランド研究所は世界初の軍事シンクタンクであり、冷戦下の戦略研究のためにゲーム理論の研究が熱心に行われていた。1950年代のアメリカでは軍からの資金で学際的な研究をすることが盛んであった。ドイッチも海軍の委託プロジェクトを受けて、ゲーム理論の専門家との交流を通して、紛争解決や平和構築のような研究に行き着いた。ドイッチの研究は社会心理学者の間で広まるというよりは、基本的には行動科学関係者の間で、囚人のジレンマを用いた信頼研究、および「主観的確率としての信頼」として広まった。

3.2.6 その後の展開

その後、行動科学運動は、さまざまな分野に影響を残すこととなった。社会学では、ドイツの社会学者Niklas Luhmann（ニクラス・ルーマン）が、ジンメルとドイッチの研究に依拠して信頼論を展開した。パーソンズ社会学の批判から始まったとされるエスノメソドロジーの分野でも、最初期には信頼研究が行われたが、相互行為についての着目するようになるにつれて、信頼という言葉はあまり使われなくなった。また、1980

年代になると英国の社会学者Anthony Giddens（アンソニー・ギデンズ）らによってリスク社会論が登場し、信頼が注目されるようになったが、これは別の系譜に属する。

心理学では、社会心理学者山岸俊男の信頼論は、ルーマン信頼論を批判したものだが、ルーマンが依拠している同じ社会心理学者のドイッチの議論は無視している。山岸は安心と信頼をどちらも認知バイアスとして定義しているが、ドイッチは信頼を紛争解決の要因とみなしており、両者は信頼の異なる側面に注目していると言える。現在の社会心理学では、信頼に関するさまざまな尺度が作成されている。

政治学では、1950年代から行動科学運動の影響があり、1960年代にはパーソンズのシステム理論に基づく政治文化論が生まれるが、信頼がクローズアップされることはなかった。1990年代にはアメリカの政治学者Robert Putnam（ロバート・パットナム）によってソーシャルキャピタル論が広まり、その中で信頼研究が行われるようになったが、これは別の系譜に属する。パットナムを批判するアメリカの政治哲学者Russell Hardin（ラッセル・ハーディン）の信頼論は、「主観的確率としての信頼」を発展させたものである。

他にも、経営学は行動科学の一分野であり、経営学での信頼は主に「主観的確率としての信頼」を発展させたものとして使われている。哲学では1980年代以降、「主観的確率としての信頼」を「Reliance」とし、信頼（Trust）と区別するようになった。教育思想での信頼は、協調行動や紛争解決などとは異なる種類の信頼、「無償の愛」に近い概念として使われている。医療人類学・医療社会学での信頼は「トラスト」以外にラポール（Rapport）がある。

3.2.7 まとめ

人文・社会系の諸分野における信頼は、大まかな共通性は見いだせるが、内実はさまざまである。それでもあえて主要なキーワードを挙げると、一つは主観的確率、リスク、期待と呼ばれるもので、二つ目は誠実さ、愛、善意など、情動に関わるもの、三つ目は社会の複雑さ、社会的規範、社会制度など、社会に関わるもの、となる。

最低限の定義を考えるのであれば1番目の主観的確率に関するグループを採用するということはあり得る。しかし、実際にはこれらを複雑に組み合わせているので、われわれがトラストを考えるときにも、こういう要素が組み合わさったものをトラストと呼ぶ場合もあるということを念頭に置くべきである。囚人のジレンマもトラスト・信頼概念の中心的なものだが、それだけではないということに注意が必要である。

今後の課題としては、まず、人文・社会系の信頼研究の大半は、組織が含まれるにせよ、基本的には人と人との信頼を対象としているため、デジタル社会にそのまま適用しづらく、調整が必要だろう。また、これだけ多様な信頼・トラストを考えると、主観的確率としての信頼を中心に位置付けることは可能だが、より幅広いものがあり得るので、それらをカバーするには分野横断的な取り組みが必要になるだろう。それは、実際に囚人のジレンマに関して行動科学運動で起きたことなので、ヒントとなりうる。

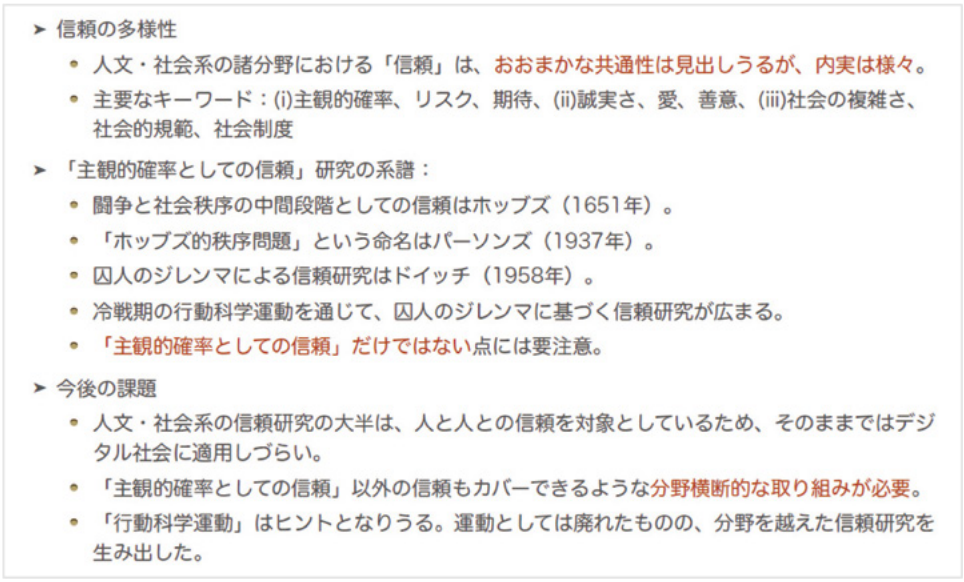


図3–2–4　　まとめ

【質疑・討議】

福島：人と人との信頼が中心で、例えばAIやシステムに対する信頼はあまり触れられていなかったとのことだが、例えばアクターネットワーク理論など、人以外も扱われつつあるといった状況はどうか。

小山：特にこの10年、20年ぐらいで機械やロボットなりに関するトラスト・信頼に関する研究がいろいろ出てきた。複数の要素が考えられる場合が多くて、主観的確率としてリスクを見積もるということに加えて、関係性を深めるというものや、社会制度のようなものを背景として相手を信じるというものといった三つの要素で考えられていることが多い。このように、人以外に関する信頼でも共通点が見いだされていると言える。

3.3 情報系での国際的動向と分野横断の必要性

村山 優子

私からは、情報科学におけるトラスト研究の国際的動向についてお話ししたい。私はもともとロンドン大学でインターネットの研究をしていて、そこからネットワークセキュリティーに入ってきたので、ネットワーク系、分散処理系、セキュリティーといった分野から見た研究事例の紹介となる。先ほど中島先生から要求工学の視点から年代別に紹介があったので、私からの説明はそれを補うものにもなる。

3.3.1 情報科学におけるトラストの研究

実は情報科学の分野でトラストの研究に本格的に取り組まれるようになったのは1990年以降だ（表3–3–1）。

表3–3–1　　情報科学におけるトラストの研究（1）

1990年より前
分散処理：ビザンチン将軍問題：ビザンチン故障等 ・Lamport, L.; Shostak, R.; Pease, M.：The Byzantine Generals Problem, ACM Transactions on Programming Languages and Systems 4（3）: 382–401, 1982. [1][2] ・分散処理環境における合意形成の研究、不正者もいる中での情報処理
ネットワークセキュリティ：BAN論理：（belief vs. truth） ・Burrows, M.; Abadi, M.; Needham, R.: A logic of Authentication, SRC, DEC report, 1989. [3][4][5] ・認証プロトコルの安全性の形式的証明 ・認証プロトコルのメッセージ交換の受信者のBeliefを検証
1990年前後：ネットワークセキュリティ分野
セキュリティ保持：認証や署名の仕組みに「トラスト」の概念を定義無しに導入 ・公開鍵をトラストするための手段を提供：Aliceさんの公開鍵は□□□、Bobさんのは※※※
公開鍵証明書の検証：Chain of Trust ・ITU（International Telecommunication Union Telecommunication）：Recommendation X.509：The Directory - Authentication framework, originally approved on 1988-11-25, superseded
公開鍵配布におけるトラスト：Web of Trust : PGP（Pretty Good Privacy） ・Zimmerman, P.: インターネットにソフトウェア公開：1991 http://www.philzimmermann.com/JA/background/index.html ・Zimmerman, P.: Pgp: Source Code and Internals, (1995/8/1) ISBN-10 : 0262240394 ISBN-13 : 978-0262240390

［1］Lamport, L.; Shostak, R.; Pease, M.（July 1982）. “The Byzantine Generals Problem”. ACM Transactions on Programming Languages and Systems 4（3）: 382-401. doi：10.1145/357172.357176

［2］佐藤一郎:「ビザンチン将軍問題」とは何か, NII Today 第69号（2015）. https://www.nii.ac.jp/today/69/4.html

［3］BURROWS, M., ABADI, M., AND NEEDHAM, R.M. A logic of authentication. Rep. 39, Digital Equipment Corporation Systems Research Center, Palo Alto, Calif., Feb. 1989 https://www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-39.pdf

［4］BURROWS, M., ABADI, M., AND NEEDHAM, R.M. A logic of authentication. ACM Transactions on Computer Systems, Volume 8, Issue 1, Feb. 1990 pp 18-36. https://doi.org/10.1145/77648.77649

［5］長谷部浩二, 岡田光弘, BAN論理からProtocol Composition Logicへ: セキュリティプロトコルの論理的検証法（<特集>数理的技法による情報セキュリティ）, 応用数理17 (4), 311-322,(2007),
https://doi.org/10.11540/bjsiam.17.4_311,
https://www.jstage.jst.go.jp/article/bjsiam/17/4/17_KJ00004805361/_article/-char/ja

表 3–3–2　　情報科学におけるトラストの研究（2）

1990年以降
セキュリティからのトラスト（CMU 関連）： ●Whitten, A.;Tygar, D.: Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0, Proc. of the 9th USENIX Security Symposium, pp.169–184, 1999［6］ ・PGPのインタフェースを例に，トラストとユーザインタフェースの考察：secureな技術もインタフェースが悪いと使われない。 ・Carnegie Mellon University → UC Berkeley ●Camp, L.J. "Design for Trust", Trust, Reputation and Security: Theories and Practice, ed. Rino Falcone, Springer-Verlang（Berlin）2003［7］ ・トラストの概念をComputer Securityの観点から提言 ・CMUで学び、Doug Tygarの下で電子商取引で学位 ・なお、CMU CyLabからセキュリティとユーザビリティの国際会議Symposium on Usable Privacy and Security（SOUPS）は始まった。https://cups.cs.cmu.edu/soups/2005/
ヒューマンインタフェース分野から ●Riegelsberger, J.; Sasse, M.A.;McCarthy, J.D.: Privacy and trust: Shiny happy people building trust?: photos on e-commerce websites and consumer trust, Proc. of CHI2003 vol. 5, no. 1 pp. 121-128 2003［8］ ・トラストの定義。好感のもてる反応は顧客の意思決定に影響 ・ロンドン大学University College London → ドイツの大学 ●Friedman, B.; Khan, P.H.;Howe, D. C.: Trust online, CACM, Vol. 43 , Issue 12, pp. 34 – 40 2000［9］ ・People trust people, not technology. ・value sensitive design（VSD）(an approach to account for human values in the design of information systems）を推進
AIの分野 ●Marsh, S.P.: Formalising trust as computational concept, Ph.D. thesis University of Stirling 1994［10］ ・トラストを-1 から+1の範囲で定量化し，初めての計算可能なトラストモデルを提案 ・トラストの国際会議iTrustを開催、その後、IFIP WG11.11（Trust Management：TM）を設立し、毎年国大会議TMを開催。 ・現在の研究：Computational RegretやComputational Forgiveness
セキュリティと最近のAI分野 ●Anderson, R.; Shumailov, I.:Situational Awareness and Adversarial Machine Learning – Robots, Manners, and Stress, 2021［11］https://www.cl.cam.ac.uk/~rja14/Papers/situational-awareness2021.pdf ・ケンブリッジのセキュリティ研究者：セキュリティ心理学、状況把握。。。 ・ロボットの敵意について→ ロボットのマナー→ ロボットのトラスト ・こちらが脅威に感じていることを、droneに知らせるすべ（protocol）無し ・意図（intent）のさまざまな示し方：犬は、吠える、うなる
電子商取引におけるトラスト ●Xiao, S.; Benbasat, I.: The formation of Trust and Distrust in Recommendation Agents in Repeated Interactions: A Process-Tracing Analysis, Proc. of ICEC 2003, pp.287-290 2003［12］ ・推薦エージェントの利用者のトラストの感情部分に言及 ●Stephens, R.T.: A framework for the identification of electronic commerce design elements that enable trust within the small hotel industry, Proc. Of ACMSE'04, pp.309 – 314 2004［13］ ・ホームページのレイアウトや画像等のデザイン要素が，電子商取引の売り手と買い手の間のトラストに影響 ●Cheskin Research and Studio Archetype/Sapient: eCommerce Trust Study, joint project, 1999［35］プロジェクトのレポート ●Kobayashi, T. & Okada H. :The Effects of Similarities to Previous Buyers on Trust and Intention to Buy from E-Commerce Stores: An Experimental Study Based on the SVS Model, Chapter 2, IT Enabled Services, Vol. 9783709114254 Springer-Verlag Wien, pp. 19-38, 2013 DOI: 10.1007/978-3-7091-1425-4_2［14］ ・主要価値類似性（SVSモデル）の電子商取引への応用

［6］Whitten, A. Tygar, D.: Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0, Proc. of the 9th USENIX Security Symposium, pp.169-184, 1999

［7］Camp, L.J. "Design for Trust", Trust, Reputation and Security: Theories and Practice, ed. Rino Falcone, Springer-Verlang（Berlin）2003

［8］Riegelsberger, J., Sasse, M.A. and McCarthy, J.D.: Privacy and trust: Shiny happy people building trust?: photos on e-commerce websites and consumer trust, Proc. of CHI2003 vol. 5, no. 1 pp. 121-128 2003

［9］Friedman, B., Khan, P.H. and Howe, D. C.: Trust online, CACM, Vol. 43 , Issue 12, pp. 34 - 40 2000

［10］Marsh, S.P.: Formalising trust as computational concept, Ph.D. thesis University of Stirling 1994

［11］Anderson, R., Shumailov, I. : Situational Awareness and Adversarial Machine Learning - Robots, Manners, and Stress, 2021 https://www.cl.cam.ac.uk/~rja14/Papers/situational-awareness2021.pdf

［12］Xiao, S. and Benbasat, I.: The formation of Trust and Distrust in Recommendation Agents in Repeated Interactions: A Process-Tracing Analysis, Proc. of ICEC 2003, pp.287-290 2003

［13］Stephens, R.T.: A framework for the identification of electronic commerce design elements that enable trust within the small hotel industry, Proc. of ACMSE'04, pp.309 - 314 2004

［14］Kobayashi, T. & Okada H. : The Effects of Similarities to Previous Buyers on Trust and Intention to Buy from E-Commerce Stores: An Experimental Study Based on the SVS Model, Chapter 2, IT Enabled Services, Vol. 9783709114254 Springer-Verlag Wien, pp. 19-38, 2013 DOI: 10.1007/978-3-7091-1425-4_2 https://www.researchgate.net/publication/287235476
［35］Cheskin Research and Studio Archetype/Sapient: eCommerce Trust Study, joint project, 1999 https://files.transtutors.com/cdn/uploadassignments/80515_2_cheskinresearchtrust1999.pdf

それ以前に出てきたビザンチン将軍問題は、私のロンドン時代の研究において、ネットワーク構成のホストやルーターの位置情報はどうすれば信じることができるかという問題に直面したときに紹介されたもので思い入れが深い。これは敵と味方に分かれている状況で、味方だと思っている中にも敵がいるかもしれないときの合意形成の問題を扱う。

また、1980年代後半には、ネットワークセキュリティー分野でBAN論理の研究がある。それまで、こういう認証プロトコルについては、提案手法の主張中心の論文ばかりだったが、当該論理により、それを客観的に証明した。そこでは、トラスト自体ではないが、BeliefやTruthが扱われている。ロンドン時代に出会った哲学の研究者から「世の中にはTruthはない、Beliefがあるのみだ」と言われた。私は「IPアドレスはTruthに決まっているじゃないか」と思ったが、よく考えてみると、これは内容真実性の話につながるが、事実だと思うことを私個人がどのようにして認識しているかというと、何かのメディアを通した情報など、自分で確かめた情報ではないことも多い。そのようなことから、私の身の回りにあるものは全てBeliefであるように思える。その上でトラストとは何だろうかと考えている。

1990年代に入る頃、セキュリティー保持のために認証や署名の仕組みに、トラストの概念が定義なしに入ってきた。通信関係でX.509という、公開鍵をトラストするための手段が提供され、Aliceの公開鍵は何々だ、Bobの公開鍵は何々だという情報の証明書や、お互いにサインし合うというWeb of Trustといった概念が出てきた。ここでいうトラストは対象真正性にあたるものだと思う。

続いて1990年代から情報科学のトラストの研究が進む（表3–3–2）。特にDoug Tygar、Jean Campらによってトラストに関する概念の研究が進んだ。先ほどの中島先生の話のように、要求工学やシステムに近い方の研究では、トラストと言わずにディペンダビリティーという言葉がよく使われていたが、ヒューマンインタフェースの分野は人間寄りなので、トラストという言葉が使われ、特に好感の持てる反応が電子商取引などで顧客の意思決定に影響するという研究がある。

それから、Batya Friedmanらの論文では、トラストは人に対するもので、テクノロジーに対するものではないと主張されたけれども、今やサービスやテクノロジーに対するトラストについても皆考えるようになってきた。それから、AI分野でStephen Marshが、初めて計算可能なトラストモデルを提唱した。さらに、セキュリティーと最近のAI分野では、BAN論理のRoger Needhamの下で研究に取り組んでいたRoss Andersonが、セキュリティー心理学を考えたり、最近はドローンやロボットのトラストといった課題にも取り組んでいる。

情報科学におけるトラストは他にもいろいろな分野があるが、電子商取引の分野では、Sherrie Xiaoらが安心感、トラストの感情部分という観点から電子商取引アプリケーションの研究をしている。それから、Todd Stephensは、ホームページのレイアウトや画像などのデザイン要素が電子商取引の売り手と買い手の間のトラストに影響することを示した。Tetsuro KobayashiとHitoshi Okadaの共著で、Eコマースにおけるトラストの論文も出ている。

セキュリティー分野では、Lance Hoffmanらによるトラストモデルの論文が2006年にCACM誌に掲載された。私もそれをリファーしたモデルを発表したので、後で紹介する。表3–3–3で「安心の研究」としたところは、私自身の研究を挙げている。

さらに、東日本大震災を契機に、災害におけるトラストがいろいろ検討されている。Yuko TanakaとYasuaki SakamotoらがSNSにおける誤情報の対策としてクリティカルシンキングを提唱した。日本のNICTは、AIを使った対災害SNS情報分析システムDISAANAを開発した。これは人々のツイートから最新の災害情報を集める手段で、良い情報も悪い情報もうその情報もあるかもしれないが、見る人が判断するというタイ

プのシステムになる。

他にも人間工学分野では、1980年代から、Situation Awareness、状況認識と訳されることが多いが、私はあえて状況把握と呼ぶ。3段階モデルで状況把握を考えている。レベル1はPerceptionで、周りにどのようなものがあるかという情報収集。レベル2はComprehensionで、それらの意味するところは何かという理解。レベル3はProjectionで、理解して分かったことから1時間後、2時間後がどうなるかという予測。これはもともと米国の爆撃機がどのようにして対象のところに爆撃するかということから考えられた。それをJohn Harraldらは災害情報共有に使った。ここでもトラストが重要で、それを利用してTaro KannoとKazuo Furutaらは、相互Beliefに基づくチームの状況把握のための認知モデルを提案した。

表3-3-3　　情報科学におけるトラストの研究（3）

1990年以降
セキュリティ分野のトラストや安心 ●トラストモデル ・Hoffman, L. J., et al. : Trust beyond security: an expanded trust model, Comm. ACM, Vol. 49, No.7, pp.94–101 2006［15］: operating systemのセキュリティ、情報のassurance等の研究 ・Pradip Lamsal: Understanding Trust and Security. Technical Report, Department of Computer Science, University of Helsinki, Finland, 2001［34］: トラストモデルのサーベイ論文 ●安心の研究 ・Murayama, Y., Hikage, N., Hauser, C., Chakraborty, B. and Segawa, N.: An Anshin Model for the Evaluation of the Sense of Security, Proc. the 39th Hawaii International Conference on System Science (HICSS'06), Vol.8, p.205a 2006［36］ ・日景奈津子,カール・ハウザー,村山優子: 情報セキュリティ技術に対する安心感構造に関する統計的検討, 情報処理学会論文誌,48（9）3193–3203 2007［30］ ・西岡大, 村山優子：オンラインショッピング時の安心感における情報セキュリティ技術に関する安全とユーザ属性との関係, 情報処理学会論文誌,55（9）2168–2176 2014［31］
災害対応におけるトラスト ●Murayama, Y., et al.: Trust Issues in Disaster Communication, Proc. the 46th Hawaiian International Conference on System Sciences (HICSS–46), pp.335–342 2013［16］ ・東日本大震災における経験から、トラストの必要性 ●Tanaka, Y., Sakamoto, Y., Honda, H.：The Impact of Posting URLs in Disaster–related Tweets on Rumor Spreading Behavior, Proc. the 47th Hawaii International Conference on System Sciences (HICCS–47), pp. 520–529, 2014［17］ ・SNSにおける誤情報の対策としてCritical Thinkingを提唱 ●大竹清敬, 他：対災害SNS情報分析システムDISAANA, NICT NEWS, no. 452, pp. 8–9 2015［18］ ・今現在のTwitterへの投稿をリアルタイムに分析 ・発生している災害に関する問題・トラブルを自動的に抽出 ・誤報対策：矛盾する投稿をどちらも提示
人間工学（ergonomics）分野 ●状況把握（Situation Awareness） ●Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32–64 (1995).［19］ ・各個人の状況把握の3段階モデルを提唱 ・SA Level 1: Perception of the Elements in the Environment（情報収集） ・SA Level 2: Comprehension of the Situation（理解） ・SA Level 3: Projection of Future Status（予測）
●Harrald, J., Jefferson, T.: Shared Situational Awareness in Emergency Management Mitigation and Response. Proc. HICSS–40, pp. 23–23, IEEE, Waikoloa, HI, USA (2007).［20］ ・災害対応における状況把握の共有のために必要なこと： ・技術ができること：離れた場所にいる意思決定者に同じ情報を届ける ・その情報をまとめ、構造化し、理解するための共通手法がある ・重要な意思決定する際、制度、文化、体験の基盤を共有し、理解した意味合いを知識とする ●・Kanno, T, Furuta, K, and Kitahara, Y.: A Model of Team Cognition based on Mutual Beliefs. Theoretical Issues in Ergonomics Science, 14（1), 38–52 2013［21］ ・相互の信用（belief）に基づくチームの状況把握のための認知モデルを提案

［15］Hoffman, L. J., et al. : Trust beyond security: an expanded trust model, Comm. ACM, Vol. 49, No.7, pp.94-101 2006
［16］Murayama, Y., et al.: Trust Issues in Disaster Communication, Proc. the 46th Hawaiian International Conference on System Sciences (HICSS-46), pp.335-342 2013
［17］Tanaka, Y., Sakamoto, Y., Honda, H.:The Impact of Posting URLs in Disaster-related Tweets on Rumor Spreading Behavior, Proc. the 47th Hawaii International Conference on System Sciences (HICCS-47), pp. 520–529, 2014
［18］大竹清敬, 他：対災害SNS情報分析システムDISAANA, NICT NEWS, no. 452, pp. 8-9 2015
［19］Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32-64 (1995).
［20］Harrald, J., Jefferson, T.: Shared Situational Awareness in Emergency Management Mitigation and Response. Proc. HICSS-40, pp. 23-23, IEEE, Waikoloa, HI, USA (2007). https://ieeexplore.ieee.org/document/4076416?reload=true&arnumber=4076416
［21］Kanno, T, Furuta, K, and Kitahara, Y.: A Model of Team Cognition based on Mutual Beliefs. Theoretical Issues in Ergonomics Science, 14(1), 38–52 2013 doi:10.1080/1464536X.2011.573010
［30］Lewis, J. D.; Weigert, A. :Trust as a Social Reality, Social Forces 63(4) 967-985 Oxford University Press 1985 https://doi.org/10.2307/2578601 https://www.jstor.org/stable/2578601
［31］日景奈津子，カール・ハウザー，村山優子：情報セキュリティ技術に対する安心感構造に関する統計的検討，情報処理学会論文誌，Vol.48, No.9，pp.3193-3203 2007
［34］Pradip Lamsal. Understanding Trust and Security. Technical Report,Department of Computer Science, University of Helsinki, Finland, 2001. http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=B347CB60E1EADBBB2E59A9A9EAB4C7A8?doi=10.1.1.17.7843
［36］Murayama, Y., Hikage, N., Hauser, C., Chakraborty, B. and Segawa, N.: An Anshin Model for the Evaluation of the Sense of Security, Proc. the 39th Hawaii International Conference on System Science (HICSS'06), Vol.8, p.205a (2006). https://ieeexplore.ieee.org/document/1579709

3.3.2 安心の研究とトラスト

コンピューターセキュリティーでは、技術さえ提供すればユーザーは安心するんだという考えの研究が多かったが、本当に安心するのかと思って、私は何年か前に研究を始めた。セキュリティーもトラストも片仮名言葉なので、日本語として意味が取りづらい。セキュリティーの技術面は考慮されていたが、その目的とする安心感についてはあまり考えられていなかったので、研究することにした。

図3–3–1は、安全と危険、安心と不安という2軸で表した。この図に示すように、セキュリティーの研究というのは、大体安全なシステムを皆が安心するようにすることと、もう一つ、安全でないシステムでは安心しないということも目標になると思う。

図3–3–2のように、関連するいろいろな言葉がある。セキュリティー（Security）は、不正を働く人がいるという心的な脅威に対する。セーフティー（Safety）は、身体的な脅威や健康上の脅威に対するもので、通常は自然災害など故意でないものを考えている。ナショナルセキュリティー（国家安全保障）という場合は、セーフティーも関わってくる。リライアビリティー（Reliability）は、物理的な故障率など物理的な脅威に対して考えられた。その後、いろいろなものをネットワーク化したソフトウェアも含むようになり、ディペンダビリティー（Dependability）という言葉も使われるようになった。ディペンダビリティーと同じような概念として、トラスト（Trust）は、心的な脅威に対するもので、故意のものも故意でないものも含む。

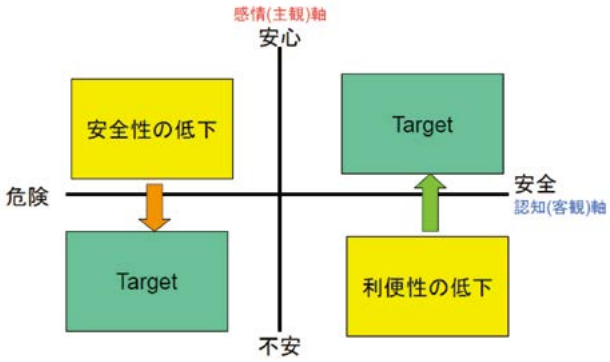


図3–3–1　安心と安全

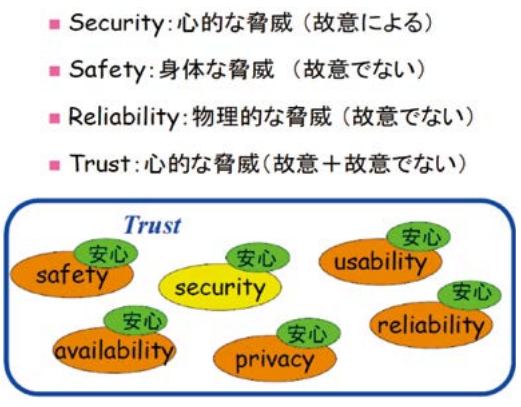


図3–3–2　安心やトラストに関連する概念

表 3–3–4　　心理学・社会学・経済学等におけるトラストの研究

安心とは ●Emotional part of Trust/Dependability
安心とトラストの差異 [22] ●安心な環境：自分から搾取する要因が相手に存在しない。社会的不確実性の低い環境ではトラストは必要ない。 ●トラストが必要な環境：自分から搾取する要因が相手に存在。社会的不確実性の高い環境。
社会心理学・組織心理学における安心の研究：リスクコミュニケーション ●知識のある安心 ●知識の無い安心⇒能動的な知識取得が必要 ・専門家は，技術的なリスクを減少させる努力だけでなく，情報提供も必要であると提案 ・吉川,白戸,藤井,竹村：技術的安全と社会的安心，社会技術研究論文集,vol.1，pp.1-8 2003 [23]
情報科学（ヒューマンインタフェース）におけるTrustの定義 [8] ●Trustはriskが存在するときのみ必要 ・Riskはさまざまな分野により定義や意味が異なる（経済学，社会学，リスクマネージメント） ・Riskはan adverse outcome（不利な結果）の可能性を表す、同義：uncertainty ・Trustはan adverse outcomeの可能性があるならば必要
トラストの研究：社会学，心理学，経済学など学際的 ●Deutsch, M.（社会心理学）[24][25]： ・個人間のトラストを定義し,その後,他者に期待することをトラストにおけるconfidenceとして紹介 ●Gambetta, D.（社会科学）[26]： ・他者の行動が自分に対し好意的かどうかという個人の主観的な確率からトラストを定義
リスク心理学
トラスト（信頼）の非対称性原理 ●信頼を得るには,肯定的実績の積み重ねが必要であるが、信頼の失墜は遥かに容易！ ●Paul Slovic[27]： 1. 信頼を崩す出来事は伝え易い 2. 否定的な事柄は,肯定的なものより,信頼評価への影響が大 3. 否定的な事実は,肯定的なものより,一般化されやすく,危険性を主張する場合の論拠に使われ易い 4. 信頼の欠如はさらに信頼を低下させるように,以降の情報処理の枠組みをかたち作ってしまう：不信の払拭は困難
信頼（トラスト）構築のための要素 ●トラスト構築の3要素：Ability/Competence（能力）、Benevolence（善意）、Integrity（誠実） ●主要価値類似性（salient value similarity: SVS model）[28] [29] [14] ・問題に関心がある場合に信頼を得るためにはメッセージ発信者との間に問題への表象の共有が条件と考えられるモデル。問題への表象とは、心の中に形成されるその問題に関するイメージのこと。 ・それに対し、従来の信頼性獲得モデルとされてきた、伝統的信頼モデルは、能力や公正性（まじめさや立場の公正さ、リスクにさらされる人たちへの思いやり）を重んじるもの」 ・上記は以下のサイトより引用：国立保健医療科学院：保健福祉職員向け原子力災害後の放射線学習サイト https://ndrecovery.niph.go.jp/?record_id=693&mode=index
トラスト感情部分の要因 ●Emotional trust：社会心理学 [30]、電子商取引 [12] ●情報科学における安心の研究 [31][32] ●精緻化見込みモデル [29][33]

[22] 山岸俊男：「信頼の構造こころと社会の進化ゲーム」東京大学出版会1998

[23] 吉川,白戸,藤井,竹村：技術的安全と社会的安心，社会技術研究論文集,vol.1，pp.1-8 2003

[24] Deutsh, M.: The effect of motivational orientation upon trust and suspicion, Human Relation 13 pp. 123-139 1960

[25] Deutsh, M.: The resolution of conflict, Yale University Press, 1973

[26] Gambetta, D.: Can we trust trust?, in Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford, chapter 13, pp. 213-237, originally published from Basil Blackwell 1988 Available online from the following site: http://www.sociology.ox.ac.uk/papers/gambetta213-237.pdf

[27] Slovic, P. :Perceived risk, trust, and democracy. Risk Analysis, 13, 675-682 1993

[28] Earle, T. C. & Cvetkovich, G.: Social trust: Toward a cosmopolitan society. Westport, CT: Praeger Press 1995

[29] 中谷内一也：安全．でも，安心できない―信頼をめぐる心理学, ちくま新書 ISBN: 978-4-480-06449-3 2008

[32] 西岡大, 村山優子：オンラインショッピング時の安心感における情報セキュリティ技術に関する安全とユーザ属性との関係, 情報処理学会論文誌, Vol.55, No.9，pp.2168-2176 2014

[33] Petty, R. E., & Cacioppo, J. T. :Attitudes and persuasion: Classic and contemporary approaches. Dubuque, IA: William C. Brown 1981

前述のLance HoffmanのモデルやJean Campのモデルによると、トラストは、セーフティー、セキュリティー、ユーザビリティー、アベイラビリティー、プライバシー、リライアビリティーなど皆含む。そのそれぞれに安心があり、日本でいう安心という言葉は、それら全てを含むのではないかと思っている。

3.3.3 多分野にまたがるトラスト研究

社会心理学や組織心理学の分野では、このような方面について国内では、山岸俊男、吉川肇子らによって随分研究されてきた。海外では、Jens Riegelsberger、Diego Gambetta、Morton Deutsch、リスク心理学でいうとPaul Slovicなどが知られている。トラストのモデルとして、中島先生の講演でも出てきたABIモデルはトラスト構築の3要素を示している。主要価値類似性のSVSモデルは、相手が自分と同じ痛みあるいは夢を共有できていると思うと、そういう相手をもっとトラストするというモデルである。また、トラストには感情部分があると考えている。精緻化見込みモデルは、いろいろな情報を受け取ったとき、動機付けが高かったり、能力がある人は自分で判断するが、そうではないと、情報そのものではなく、それを誰が言ったか次第で信用するのではないかというモデルである。

このようにトラストの研究分野は幅広く、分野横断的に取り組まないと極められないだろう思っている。トラストを定義せずに、飾りに付けているようなものも世の中では見られるが、できるだけ明確に定義して、分野横断で多角的に研究していくことが大切だと考えている。

3.4 ステークホルダー関与、医療分野での実践事例

山本 ベバリー アン

ステークホルダー関与によるトラスト形成について議論したい。今までの話題と異なってかなり具体的になるが、特に医療分野における患者、研究者、臨床研究者の間での関与と実践事例について話す。

3.4.1 医療におけるステークホルダー関与の意義と広がり

背景として、HIV/エイズが広がった1980年代の米国で、医療分野では恐らく初めて患者と研究者が話し合いさまざまな問題を解決するようになった。当時のことが2022年12月31日のニューヨークタイムズに掲載されている[24]。当初HIV/エイズに関するガイダンスがなく、薬の治験も非常に時間がかかる中で、患者活動が活発化し研究者と衝突したが、一緒に話し合えば良い結果が出た。その後、国立衛生研究所（NIH）が関係するほとんどのエイズ研究の委員会に患者代表が参画するようになった。こうして医療における患者関与モデルが誕生した。

その記事によると、最近の新型コロナウイルスのまん延初期にも、まだワクチンも薬もないという状況において、かつての活動家でHIV/エイズ生存者が経験を生かし相談相手になったようだ。つまり、患者側と研究者、医療従事者側が直接話すなどして共同すると良い結果になる。

ステークホルダーという言葉は日本語としてあまりなじみがないが、主な定義としては「直接的または間接的に影響を受ける利害関係者」のことである。医療分野のステークホルダー（図3–4–1）はさまざま存在するが、私たちの研究では特に患者・家族、医師・看護師、医療行政の人々が一緒に話し合うことを主な研究対象としている。

図3–4–1　医療のステークホルダー

こうした患者・市民の関与は英語でPatient and Public Involvement（PPI）、最近はPatient and Public Involvement and Engagement（PPIE）とも呼ばれる。前述のようにHIV/エイズの広がった1980年代から患者と研究者が話し合うことの重要性が理解され、さまざまな委員会に患者代表が参画するよ

24 “Anthony Fauci Quietly Shocked Us All” Dec. 31, 2022, The New York Times
https://www.nytimes.com/2022/12/31/opinion/anthony-fauci-hiv-aids-act-up.html

うになった。そして1990年代から英米を中心に希少疾患、慢性疾患の患者の関与が活発になった。臨床治験などの医学研究プロセスに患者が関与・参画するPPIによって、直接患者の声が入り研究の価値が上がると期待されて広まった。

患者側、特に患者団体の主導で研究助成を行う財団、バイオバンク、レジストリーなどを設立した事例は少なくない。また、行政側もPPIの利点をよく理解し、英国 国立衛生研究所（NIHR）、米国 患者中心のアウトカム研究所（PCORI）など大きな機関が活発にPPIを実施している。

2017年には英国 NIHRがPPIを“Research ‘with’ or ‘by’ members of the public rather than ‘to’, ‘about’ or ‘for’ them”（患者・市民のために、または患者・市民について研究が行われることではなく、患者・市民と共に、または患者・市民によって研究が行われること）と定義した[25]。

近年は日本も日本医療研究開発機構（AMED）などを介してPPIを推進している。2019年、AMEDが「PPIガイドブック」[26]を発行、医薬品医療機器総合機構（PMDA）が患者参画検討ワーキンググループを発足（のち2021年「患者参画ガイダンス」を公開）[27]、厚生労働省が全ゲノム解析等実行計画（第1版）において「全ゲノム解析等を推進するにあたり、患者・市民参画の仕組みを設けるなどELSIへの対応ができる体制の在り方等について検討する。」と記載[28]、さらに一般社団法人 PPI JAPAN（PPIコンソーシアム）が設立された。2019年は日本のPPI元年と言ってもよいだろう。

PPIにおける患者の役割は多様である。治験の委員会ほか、さまざまな運営委員会、アドバイザリー委員会などに参加する患者代表として。それらの活動促進グループ（患者グループ）として。また最近は、研究チームに入っていろいろなアドバイザリーの役割を果たすこともあり、ペイシェント・オーサーという言葉も生まれた。私たちが調査したところ、実際にこの2～3年間でさまざまな医学論文にペイシェント・オーサーが記載されている。学術雑誌がどのような推奨、推薦をしているかについても調査したところ、例えばブリティッシュ・メディカル・ジャーナル（BMJ）では2014年から“Partnering with Patients”のプログラム、戦略があり、研究のアウトプットの段階でPPIがどのように反映されているかのチェックシートもある[29]。

3.4.2 社会の技術基盤の変化による社会的な関係の変化

では、PPIのようなステークホルダー関与が実社会でどのように役立つだろうか。経団連が図示したように、人類社会は技術の変化とともに社会関係も変わってきた[30]。そして今はSociety4.0（情報社会）からAI、IoT、ブロックチェーンなどのデジタル革新によって社会が大きく変化しつつある。対人経験の仕方が直接対面よりもスマートフォン、Zoomなどを介して人間関係を感じるようになった。

日本政府の「Society5.0」という概念は、英国や米国で言われる「Good Society」と似ているように思う。技術はもっと社会のため、みんなのための良いものにならないといけない。そのために倫理的な影響評価の理念が大事ではないか。どのような人に、どのような影響があるか分からないと、Good Societyは作れないだろうという考え方だ。

新興技術は、根本的な新規性、比較的急速な成長、一貫性と概念の自律性、顕著な影響と不確実性、そ

25 NIHR “What is public involvement in research?” https://www.nihr.ac.uk/patients-carers-and-the-public/i-want-to-help-with-research/

26 AMED「患者・市民参画（PPI）ガイドブック」https://www.amed.go.jp/ppi/guidebook.html

27 PMDA「患者参画検討WG」https://www.pmda.go.jp/rs-std-jp/cross-sectional-project/0020.html

28 厚生労働省「全ゲノム解析等実行計画（第1版）」https://www.mhlw.go.jp/content/10601000/000579016.pdf

29 BMJ “Patient and public partnership” https://authors.bmj.com/policies/patient-public-partnership/

30 日本経済団体連合会「Society 5.0 －ともに創造する未来－」http://www.keidanren.or.jp/policy/society5.0.html?v=s

して曖昧さによって定義されている[31]。ヘルスケア分野の新興技術の中で私たちが注目しているAIとデータリポジトリーについては、特にCoherence（一貫性）とConceptual Autonomy（概念の自律性）が重要だ。AI技術を一般の人が理解できるか、一般の医療従事者でも理解できるか、十分な情報があるか、などを研究プロジェクトにおいて検討しているところだ。

新しいテクノロジーで社会基盤が変わると、人は新しい方法で人間関係や社会関係を経験することになる。新しい技術をトラストできるかは重要な問題だ。信頼できないと社会に受け入れることは難しい。

トラストにはさまざまな定義があるが、議論のために二つの側面で考えてみよう。一つは新しい技術について受け取る情報への信頼。もう一つは新しい技術を開発している人々への信頼。PPIではそれらの両方が見られると思う。

AI技術へのステークホルダー関与の例として、2021年の国連教育科学文化機関（UNESCO）による“Recommendation on the Ethics of Artificial Intelligence”（AI倫理に関する勧告）[32]が挙げられる。勧告では「Ethical Impact Assessment：EIA（倫理影響評価）を促進する必要があり、それにはステークホルダーの関与、参画が必要だ」と強調している。どのようなステークホルダーがどのように関与、参画するべきか、私たちも検討している。

3.4.3 事例から見たデジタル社会におけるPPIの期待と課題

私たちはPPIに関して三つの研究をしている。中心メンバーは大阪大学大学院医学系研究科の加藤和人教授と私、そして古結敦士（コゲツ アツシ）助教だが、プロジェクトによってメンバーは多少異なり、もっといろいろな研究者が入っている。

事例①

RUDY JAPAN[33]は2017年12月に開始した。 RUDYは難病・希少疾患をかかえる患者の症状や日常生活に関する情報を収集し、疾患への理解を進める医学研究プロジェクトである。英国のオックスフォード大学で始まったRUDYの日本版を大阪大学大学院医学系研究科が事務局となって立ち上げ、英国のプロジェクトと連携しながら研究を進めている。

RUDY JAPANには二つの側面がある。一つは希少疾患を対象とするPPIを日本で実践すること。日本で本当にPPIは可能かどうかの実践である。もう一つはICTを使って患者参加の医療研究を促進すること。希少疾患に関わるクオリティー・オブ・ライフ（QOL）とペイシェント・レポート・アウトカムの調査によって、患者の主体的な立場から研究を進める。開始から5年後の現在、90人の患者さんがこのプラットフォームに関わってデータを創出している。

PPIには運営委員会があり、主に患者と研究者がプロジェクトの方向性などを決める。臨床研究のための調査方法も、現在は三つの疾患を対象に患者と研究者が一緒に進めている。例えば遺伝性血管性浮腫（HAE）の発作の記録ツールが患者と研究者により作成された。

PPIにおいて、運営委員会に参加するのはかなり深いレベルの関与、参画だが、データを提供するのはもう少し浅いレベルかと思う。データ提供に関しては、参加者が自身のデータの利用や連絡方法について選択、

31 Nelson and Gorichanaz, 2019 “Trust as an ethical value in emerging technology governance: The case of drone regulation”
https://doi.org/10.1016/j.techsoc.2019.04.007

32 UNESCO “Recommendation on the Ethics of Artificial Intelligence”
https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

33 RUDY JAPAN　https://rudy.hosp.med.osaka-u.ac.jp/

変更できるような新しい同意モデル「ダイナミックコンセント」を実装した。

RUDY JAPANは研究者と患者の距離が近いことによって信頼関係の形成が可能になる、新しい医学研究だと思う。

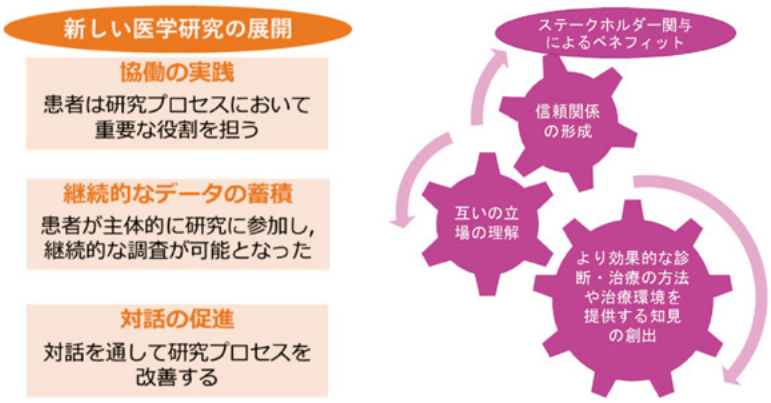


図3–4–2　　新しい医学研究の展開とステークホルダー関与によるベネフィット

事例②

コモンズプロジェクト（JST RISTEX 科学技術イノベーション政策のため科学 研究開発プログラム「医学・医療のためのICTを用いたエビデンス創出コモンズの形成と政策への応用」）[34]は政策形成のレベルでPPIが可能かどうかを研究している。どの研究に助成金を出すなどの優先がなされるかの意思決定に、最初の段階からステークホルダーとして患者の意見を入れるべきではないかという発想でスタートした。

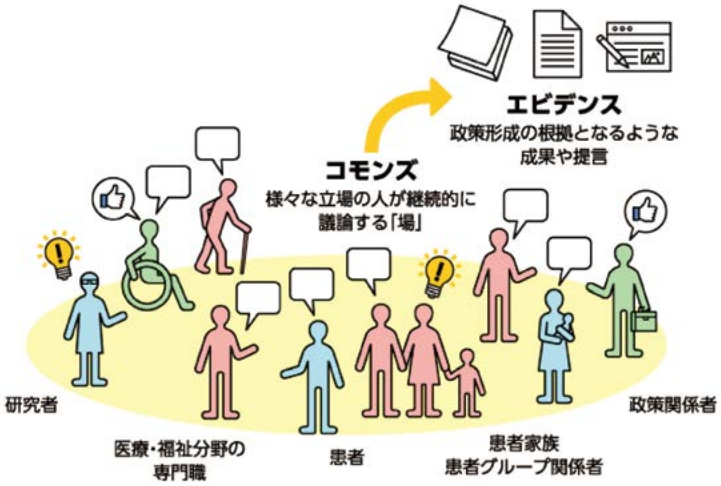


図3–4–3　　コモンズプロジェクトにおける「コモンズ」と「エビデンス」

34　コモンズプロジェクト https://www.med.osaka-u.ac.jp/pub/eth/research-project/

ステークホルダーはワークショップを全てオンラインで継続的に行い、一緒に検討を進めてきた。10疾患の患者・患者関係者21人、研究者17人、行政経験者5人の計43人が通しで定期的にワークショップへ参加した。2019年から2022年で段階的に研究を深め、どのような課題が優先で、研究を進めるために何が大事なのか議論し、最後に提言書を作成した。

研究費の配分のような政策形成に初期段階から患者の声を入れることを考えるプロジェクトは、日本で初めてだったのではないか。

事例③

AIDE（エイド）プロジェクト（JST RISTEX 研究開発領域「人と情報のエコシステム」研究開発課題「ヘルスケアにおけるAIの利益を全ての人々にもたらすための市民と専門家の関与による持続可能なプラットフォームの設計」、英国 UKRI Project Reference ES/T007214/1）[35]は医療のAI活用に対するステークホルダーの意見に関する日英共同研究だ。今回はステークホルダーの範囲を広げ、患者、市民、そして一般の医師、看護師、病院スタッフといった医療従事者も参加している。

AIDEプロジェクトは七つのワーク・パッケージ（WP）に分かれている（図3–4–4）。WP2「市民と患者の関与パネル（PPIP）」を中心に据え、どのWPでもステークホルダーの意見を聞いてプロジェクトを修正する。WP5「ステークホルダーに対するフォーカスグループ研究」とWP6「ステートメントリストの生成と評価」は、より広いステークホルダーを対象に意見交換や情報提供をしている。それぞれのWPに対してPPIPの役割があり、3カ月に1回、2時間程度会って深く話し合いさまざまな情報・意見交換をしている。

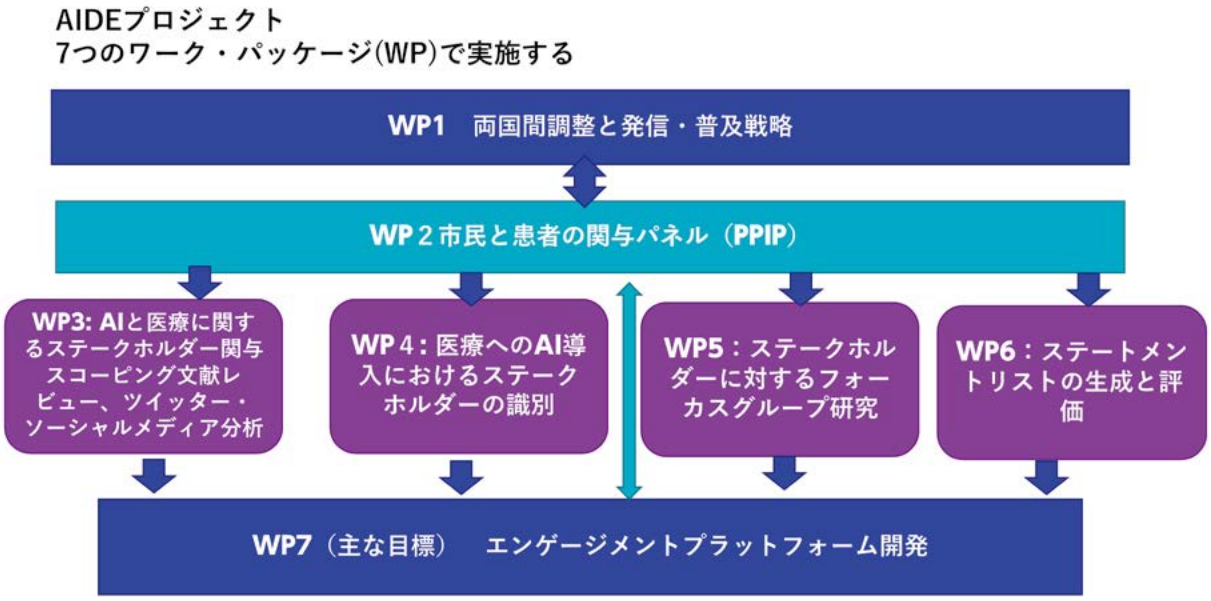


図3–4–4　AIDEプロジェクト 七つのワーク・パッケージ

より広いステークホルダー関与については、患者・市民・医療従事者・開発者のフォーカスグループ研究（WP5）で進める。ヘルスケアAIに関する現実的なシナリオを三つ作って提供し、それに対する期待と懸念、

35　AIDEプロジェクト https://aide.osaka.jp/

その考え方、信頼できるかの判断に必要な情報などについて、フォーカスグループと話し合い、結果をまとめている。現在50人以上のステークホルダーがフォーカスグループに参加している。

3.4.4 医療の例から見てきたデジタル社会における新たなトラスト形成

AIDEプロジェクトのPPIP（WP2）は深い参画モデルだ。 PPIPメンバーの依頼で、もっとAIのことを理解するトレーニングも行った。開始から3年たった現段階では、参加者が皆医療AI技術についてかなり深い知識を持ち、懸念や期待について活発に意見を出すようになった。 PPIPメンバー内でも、研究者とPPIPの間でもトラストが形成でき、かなり深い話までできるようになった。

フォーカスグループ（WP5）はより浅い関与だ。1回しか会わないし、こちらから提供する情報も限定してある。それでも対話の中で感動して、もっとこのプロジェクトに参加したいというメンバーも何人かいた。思うに、深い関与でも浅い関与でも、プロジェクトに接触、関与すること自体が信頼関係構築の基礎となるのではないか。

このように、ステークホルダー関与は信頼関係を構築する方法の一つといえる。ステークホルダーに情報を発信するだけでなく、ステークホルダーが参画することによって新しいデータ、アイデアといった情報を作り出す。参加者はその過程で学習し、話し合うための自信と力が向上する。

なおこれらの研究は実に多様な研究者が参画しており、今回の話題提供に大きく寄与していることを申し添えたい。

【質疑・討議】

福島：PPIPなど実際にかなり深く関与した人たちと医療関係者の間でのトラストが強くなるのはよく分かるが、さらにその周りの人たちもトラストできるようになるといった波及効果もさまざまに出てきているのだろうか。

山本：フォーカスグループは浅い関与とはいえ、医療従事者の場合90分間、患者の場合2時間、三つのシナリオについて話し合う。AI技術とその影響について初めて真面目に話し合い、そこからさまざまな意見が出てくる。トラストできるには情報が大事といわれるが、安全な場で話し合ってこそ情報が相互に流れトラスト形成が可能になるのではないか。PPIPでは医療AIを開発する研究者と直接話す機会があり、開発者へのトラストも形成されていると思う。またPPIPのメンバーは患者団体の代表である人も多く、プロジェクトでの経験をまず自分の患者団体に持ち帰るので、トラストが広がるのではないか。

4 ｜ 総合討議

4.1 総括コメント

中川 裕志

4.1.1 Trusted Webについて

クロサカ先生の話に関連するところから始めたい。Trusted Webの場合、トラストしているのは情報を送受する相手、それは個人だったり、グループだったり、組織だったりすると思うが、その相手の真正性という理解でよいのか。または、トラストしているのは送受される情報内容なのか。内容となると、ブロックチェーンを使えばトラストできるのか。ブロックチェーンを使うことでトラストできるかもしれないが、情報内容をブロックチェーンに乗せることはブロックチェーンの構造からして相当大変だろう。似た話としてNFTがあるが、NFTはメタ情報の真正性をブロックチェーンで担保しているのであって、内容については関知していない。電子透かしを併用するという話も聞いたことがあるが、それで完全なのか。さらに、そこまで複雑なことをやってトラストする価値がある情報とは何なのか、ということを考えることも重要なのではないか。 OSI7層レベルのどこかの層にトラストの仕組みを持っていくにしても、実装は相当大変だという気がする。情報の重要性に応じてトラストの重さ、軽さを調整するようなトラストの仕組みを考えることはできないのか聞きたい。

4.1.2 内容真実性について

次に、内容真実性については、フェイクニュースといった偽物あるいは誤った情報には社会的方策で対処する、ということを山口先生が話した。しかし、フェイクニュースの発信者は、自分が発信している情報をフェイクニュースだと思っていないことが多いので、そういった方策は難しいのではないか。以前、アノンと思われるやつらとやり合ったことがあるが、議論にならない。 Chris Bail（クリス・ベイル）の本を読んだら、フェイスニュースを流すやからをネットワーク上で孤立させる作戦がいいのではないか、ということが書いてあった。この意見には反対の方も賛成の方もいるようなので、山口先生の意見を聞きたい。

モノの真正性は、結局、モノを記述した情報をトラストするのかどうかということに行き着くのではないか。ウェブ上ではそういうことにしかならないと思う。

システムの真正性もある。これについては中島先生が振る舞いの予測可能性という意味でのトラストをいろいろ説明した。そうだろうなとは思うのだが、AIは使っているうちにシステムの予測可能性が変化してしまうという問題があるので、AIの予測可能性は非常に難しい。AIのトラストについては、Trustworthy AIとしてAI白書、AI規制法案（AI Act）といった法律で押さえつけるようなやり方で、ブリュッセル効果を狙ってEUは非常に高圧的にやる。そのやり方に便乗して果たしていいのだろうかと、大変疑問に思っている。

4.1.3 人・アバターのトラスト

非常に重要な問題として、人間をトラストできるかという問題がある。この問題を考えるにあたっては、そもそも人間とは何かという問題を考えなければいけない。人間は日々、変化しているので、何をもって真正性とするか悩ましい。生物学的実体である人間が、本人かどうか、という問題なのだが、人間の健康状態などは時間に応じて大きく変化してしまう。情報的実体としての人間という考え方もあるが、これも変わる。例えば不変な部分としてDNAとか生年月日はあるが、日々追加されていく情報、例えば年齢などは増える。ある

いは、心理的実体としての人間ということもある。これは人間の考え方で人間を捉える話だが、物の考え方は時々刻々変わるし、右翼が左翼に転向するようなこともある。人間とは何かという問題は非常に難しい。人間の捉え方で、人間をトラストする側面が変わり、理論とか技術とか制度が変わってきてしまうということである。

さらに、最近は人間だけを相手にすればいいというわけでもなくなっていて、人間の代理のAIエージェントあるいはアバターといったものがある。アバターになると乗っ取りや、なりすましがあってさらに難しくなる。イメージにすると図4–1のような状況になる。本人がいて他者エンティティーがインターネット越しにいて、間にサイバネティックアバター（CA）と呼ばれるアバターがいて、アバターは外見が違ったりする。アバターは、CAと他者との間にトラストがないと使えないし、本人とCAとの間にもトラストがないと使えない。このような構造をしているので、結構厄介な問題を含んでいる。外見を変えるアバターはトラストできるか、ということがまず考えられる。実名ではない名前を使ったアバターを使ったり、おじさんが若い女性のアバターを使ったりすることもよくあると言われている。

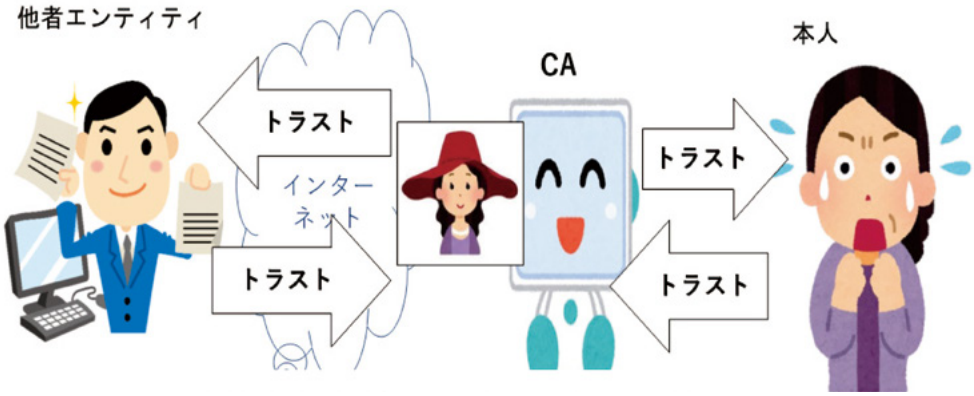


図4–1　　サイバネティックアバターのトラスト

一方で、自律的なアバターもある。個人が複数のアバターを使い始めると自律性がないと大変になるので、自律的なアバターが使われる。その自律的なアバターをトラストできるかという問題は、図4–1でいえば、他者エンティティーとCAとの間のトラストの話である。これもなかなか難しい問題で、総務省情報通信政策研究の最新号に私はこの問題について書いているし、法律的な点は石井夏生利先生が細かく分析しているので参考にしてほしい。なりすましやボットによる乗っ取りの可能性は常にあるという非常に危険な状況だが、そういった状況で本当にトラストできるのかどうかという話がある。

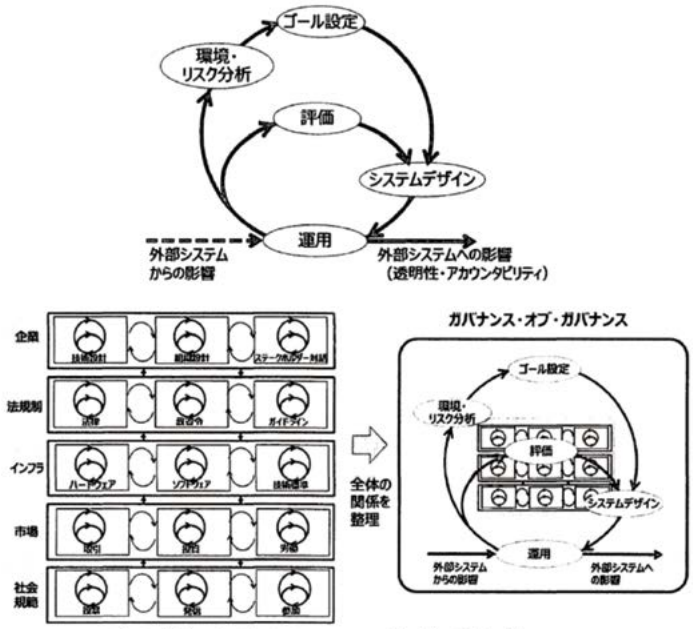
4.1.4 社会科学的観点からのトラスト

真正かどうかだけではなくて、トラストすれば被害を受けないかどうかという論点も重要である。これについては、大屋先生が関連することを話していたと思う。それから、小山先生の話にあったが、トラストしなければ得ができない、あるいは損をする、というような立て付けの議論が心理学的あるいは社会学的には最近多いように思う。

インターネット時代ではどのようなトラストを考えたらいいのか。対象の真正性、複雑さの回避、制度的な補償など、いろいろな意味でトラストを捉える考え方が出てきており、小山先生の話にあったように、非常に多様な部分に着目しなければいけないのだと思う。

話題として福島さんが簡単に触れたことだが、法制度や社会制度の仕組みにおけるトラストの問題がある。

図4–2に経済産業省の報告書から引用した図を示した。図の上側にあるカタツムリのような形は、あるシステムを運用しているときに、外部からの影響を評価して、評価結果に応じてシステムをリデザインして運用するという内側のループと、システムをコントロールしている法律が変わるとか、経済状況が変わるとかいった環境の変化やそのリスクを分析して、ゴール設定自身を変えてシステムのデザインをやり直すという外側のループの二重ループによるガバナンスを示している。図の下側はその二重ループが企業の中で、法規制、インフラ、市場といったいろいろなところで回っている様子を示している。これらの二重ループはそれぞれ独自に動きつつ、外部からの影響でお互いに関連し合う構造を持っている。



出典：Society5.0における新たなガバナンスモデル検討会「GOBERNANCEINNOVATION Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」報告書（2021年7月）

図4–2　　アジャイル・ガバナンス・サイクルとマルチステークホルダーによるアジャイル・ガバナンス

このように、法制度や社会制度の仕組みにおけるトラストの問題は、非常に自律性のあるものが動いている構造をお互いにトラストできるかどうか、という問題になるので、最近の話としては、分散的であるトラストをやや中央集権的にしようという流れの議論もあるようである。今回のシンポジウムの趣旨からは外れる懸念もあるが、アジャイル・ガバナンスは制度に対するトラストを作るときに重要なポイントになると思って紹介した。皆さまの発表を聞いて、こういった人間の問題まわりが非常に重要になってくるのだろう、ということを改めて感じた。

4.2 質疑・議論

福島：全体を振り返りつつ、多面的にさまざまな問題の指摘をしていただいたので、どう議論を進めるか悩ましいが、チャットに書き込んでいただいた神嶌さんと伊藤さんからの質問が前半の３人の発表に関する質問が多く、中川先生の問いかけにも関連しているように思うので、神嶌さんと伊藤さんに口頭で質問してもらって、多様な質問が出そろったところで、最初の３人の方々から、特に重要だと思った質問、あるいは答えておきたいと思った質問に関して答えてもらうという形でまずは進めたい。

神嶌：最も疑問に思うところは、AIがブラックボックスになっているとか、AIを導入することによって悪いことが起きるように言われている点である。私は、例えば人間が裁判で差別をやっているようなことが、AI技術によって緩和されると思っている。実際、ProPublicaでもデータ的にはそうなっている。 AIに対する現状認識が違う感じになっている点が気になった。システムを開発しているものとしては、人間よりもいいものができたら、どんどん置き換えていくことがプラスに働くと発想する。

中川：今の話に関連するが、COMPASという再犯防止システムがアメリカにあるので、そういったシステムについて社会調査したことがある。人権侵害だから反対、といったことを言う人が多いかと思っていたが、自分たちにとって役に立つものであれば許容できるのではないか、という意見の人が多かった。それから、駅といった公共の場所におけるビデオカメラも、痴漢を防止するとか、駅員への暴力を防止するとか、ある意味で自分たちにとって利益があるというものについてはあまり反対ではなく、むしろ賛成が多いくらいになっている。多くの人々は神嶌さんの言うように、自分たちに有益であれば好意的に捉えよう、という考え方をするということを、調査の結果、発見した。

伊藤：人が処理する情報は増加し続けており、それらの情報が信頼できるか否かを、人が一つ一つ判断してはいられないような社会になってきていると感じている。そのような状況の中では、人の処理能力を補うために、信頼できるか否かの判断をある程度自動化することは避けられないと思われる。

Trusted Webは、信頼するか否かをアプリケーションごとに人に判断させるのではなく、ウェブ上で機械にやらせざるを得ないという背景でのアーキテクチャーだと私は理解している。一方で、信頼プロセスが異なりうる複数のアプリケーションがあり、そのアプリケーションごとに各エンティティーが異なるポリシーを持つ場合には、エンティティー同士がお互いに信頼するか否かの判断を機械的に実施するのは非常に難しい処理に思える。このような、アプリケーションやポリシーが違う場合の信頼性の判断について、今までどおりアプリケーション層で人が判断するのか、それともウェブ層で自動的に行うのかという点に興味を持った。私の理解では、Trusted Webはウェブ層で自動処理する仕組みであるように聞こえたが、具体的な方法について検討しているものがあればご教示いただきたい。

また、フェイクニュース対策については、不正をするインセンティブを削る、対策を行うインセンティブを作り出す、不正行為の費用対効果を悪化させる、不正の対策をする側を自動化するなど、さまざまなアプローチがあるという理解をしている。これらのフェイクニュース対策は、政府機関と民間事業者が協力して取り組まなければならないと思われるが、官が得意なこと、民が得意なことは異なりうるのではないかとの印象を受けつつ聞いていた。フェイクニュース対策の中に、民が得意であろうから、ぜひ民間がやるべきではないか、という対策があればご教示いただきたい。

クロサカ：いずれも非常に有益な指摘で、Trusted Webの取り組みに今後生かしていければと思っている。まず、中川先生の指摘である、相手エンティティーは誰なのか、について答える。エンティティーは、先生がおっしゃるとおり、個人であり、グループであり、組織でありということだと思う。計算機工学的にもそうだし、契約といった社会システム上もそうだが、エンティティーは何らかの形で特定をしなければいけない。エンティティーとしてはさまざまな自然人、法人、法人の中でもさまざまな法人格があるし、個人にもいろいろなペルソナがある。これらを多様に捉えて、エンティティーとして定義できる

システムが必要だろうと考えている。計算機の上に乗っかってくるユーザーであれば、誰でもエンティティーとして定義できるようにすべきと考えている。

次に、送受される情報内容のトラストをどうするのかというのは、非常に重要な指摘だと思う。Trusted Webが、ブロックチェーンを含めて個別の技術をアーキテクチャーに指定していない、あるいは指定しなくてもいいと考えている理由の一つはそこにある。つまり、今、世の中にある、あるいは注目されている個別技術というのが必ずしもさまざまなエンティティーおよびその人たちによって構成されるトランザクションを十分にカバーできる技術になっていないだろうと思っている。

例えばブロックチェーンでいうと、ブロックチェーンに対する期待は大きいが、現実問題としてデータそのものをブロックチェーンの上に全部載せようとすると、現実的なトランザクションは不可能になる。チャレンジングに時間がかかって、10MBのデータをやりとりするのに、3時間、4時間かかるようなことが普通に起きる。では、トークンだけをブロックチェーンに載せて電子透かしで、となると、電子透かしで本当にいいのかという話が出てくる。結局、どこまでいってもそれだけでは万全ではないということである。つまり、今現在のテクノロジーは組み合わせで使うしかないので、テクノロジーそのものを厳格に定めるよりは、組み合わせによって何を達成するのかが重要なのだろうと思う。組み合わせを定義していくほうが現実的だろう。

そう考えると、三つ目の指摘である、そこまで複雑なことをする価値のある情報とは何なのかということが非常に重要になってくる。 Trusted Web全体としてコンセンサスを得ているわけではないので、私の考えということだが、情報あるいはデータそのものよりも、トランザクションの価値を定義したほうがいいのではないかと思っている。アナロジーで言うと、電子署名という技術は非常に重要な技術だが、電子契約というアプリケーションが存在しないとほとんど価値がない、マネタイズされないということがある。つまり、電子契約という手続き、トランザクションにおいて電子署名が必要だという価値がある。そう考えると、Trusted Webで相手を確認し、トラストを構成していこうとしているときに何をやりたいのか、どういうやりとりをしたいのというところから要件を定めていったほうがより具体的、現実的になる。組み合わせの問題として既存のものも含めて、例えばブロックチェーンが使えるのであれば、使えるものは使おうということになっていく。

そのときに、共通的な事項として抽出できるものが、例えばデータの形式であるとか、トランザクションの形式であるとかいうことが見つかってくれば、それはプロトコルとしてより低いレイヤーに埋め込んでいくことが可能になると思う。共通項を見つけて、できるだけ深いレイヤー、伊藤さんが指摘したウェブ層なのかということになるが、私の考えでは、第7層よりは第4層ないしは第3層あたり、つまり、TCPとかIPとか、その辺のレイヤーまで手を入れた方が本当はやりやすくなるのではないかと思う。例えばIPというのは、基本的にマシンを識別するための技術なので、識別がより明確になる、あるいはセキュアな条件が達成できるのであれば、第7層で個別に、うちの方式の電子署名を使ってくださいみたいな乱立した状況が発生するということはなくなって、よりみんなが使いやすくなるのではないかと思う。そういうわけで、どれだけ下の層に埋め込んでいくことが可能なのかにチャレンジしていくことが技術的には必要ではないだろうか。

それができると、オートメーションを考えたときに、インフラで支えている部分は自動化が可能なので、人間のポリシーによって情報であるとか、トランザクションの流通であるとかを構成することが可能になると思う。つまり、人間側で制御できる部分をできるだけ上の層において軽くすることで、よりやりやすく、テンタティブなものも採用しやすくなるようにできる、といったことになるのではないかと思っている。レイヤーに分けながら、重い部分、軽い部分を明確にして、重い部分を下層に沈めて、軽い部分を上層に上げてということを、合意形成や定義によってやれれば、レイヤーのプロトコルに埋め込めるので、そういうことをやっていきたいと考えている。

今のような話で、神嶌さんの質問にあるAIの導入のところについても、一部回答できたのではない

かと思う。

中川：同意する。

山口：中川先生の話にあった、ネットワーク上で孤立させたらどうかという話は非常に面白いやり方だと思うし、そういうアイデアが出てくるのもよく理解できる。ただ、もしそういう方向にかじを切るという話になるのであれば、なぜその人はネットワーク上で孤立させられるのかといったところの合理的な説明と透明性というのが必要になってくるだろうとも感じている。例えばツイッターのシャドウバン[36]が話題になったが、そういうことに対しては非常に批判が高まる。したがって、中長期的な施策としては難しい。実施するなら透明性が必要なのではないかと思う。

似たようなやり方として孤立まではさせないが、拡散させにくい仕組みを作るというのはあるかなと思っている。今でも既にフェイスブックやツイッターは、ファクトチェック機関と連携してチェックされたものにラベルづけしたりしている。例えばラベルづけされたものはアルゴリズムで表示されにくくするといった方法は、一つのやり方としてはあるかもしれない。もちろん、ファクトチェック自体がうそだというような主張もあるので、そう思っている人の不満は残ったままだが、客観的に説明がつくやり方としては一つの方法ではないかと思う。

また、もう一つアルゴリズムでできることでいうと、フェイクが拡散していくルートをプラットフォーム事業者は特定できるので、フェイクだということがファクトチェック機関によって明らかになったものについては、そのルートをたどって、ルート上の人たちにファクト記事を優先的に表示する、といった仕組みがあると、かなり効率的に配信できるのではないか、とも思っている。残念ながらファクトチェック記事の冒頭のフェイクを説明しているところを読んで、フェイクを信じる人というのが結構世の中にはいる。ファクトチェックが逆効果になるということもあるので、そういうことを防ぐためにもフェイクが表示された人に対して、ファクトチェックのニュースが上のほうにくるようなやり方があるかもしれないと考えている。

もう一つ、民間企業で何ができるかという話があったと思う。そこに関しては、今すぐいろいろ思いつくわけではないが、私が民間企業でフォーカスしているのはメディア企業である。メディアの役割は私の発表でとうとうと語った部分であり、民間としては非常に大きいところではないかと思っている。

メディア以外では技術的な面での貢献があると思っている。一つ大きいものとしてはディープフェイク検証である。発表の中でも言ったが、ディープフェイク技術が民主化してしまっている。 AIというのはいい面もいっぱいあるだろうという話があったように、すばらしいツールではあるが、同時に悪用できてしまうというリスクがある。典型的な例がディープフェイク画像である。 AIが勝手に画像を作ってくれるサービスを使えば、スライド資料を作るのが楽になるといった、いろいろな活用方法がある。ただ、それを悪用して誤った画像とか動画を作って、悪意を持って拡散させるという人はこれから絶対に出てくるし、最初に画像や動画を作った人には悪意があるけれども、拡散する人には悪意がなかったりする。フェイクだと気づかずに拡散してしまったりする。そういったことは防がなければならないので、ディープフェイクの検証が気軽にできる技術の開発、そして実装ということが民間に期待されている大きな点ではないかと思っている。

中島：中川先生、神嶌さん、お二人ともAIの良いところを考えて使っていくという話だったと思う。確かに私もそう思う。 Trustworthy AIといった話が出てくるのは、AIを使う側に良いものなのか、悪いものなのか分からない不安がある、だから、トラストを醸成したいとなる。そうすると、もともとあった機械

36 Twitter運営が不適切だと考えることをユーザーが行ったときに、ユーザーのアクティビティが第三者から見えにくくするようにTwitter社がユーザーに知らせずに行う操作。ユーザーは自分がその状態にあることに気づきにくいので、ShadowBanと呼ばれる。

学習やAIの良い部分を保ちたい、あるいはより良くしていきたい、ということで変化すると考えることができると思うのだが、その変化をどのように説明して安心してもらうのかが気になる。使う側からすると、トラストできる開発者が作ったものであれば使ってもよいということになるだろう。そうすると、開発者の立場からは自分をトラストしてほしい、どのようにすればトラストしてもらえるか、と考える。こういう条件を満たせばトラストしてもらえるということを見通しよく項目立てできれば、よりAI技術を使うことに関して不安がなくなるのではないかなと、そんなふうに思っている。

福島：次に、今議論されたことに関してぜひ言っておきたいこととか、中川先生の話の後半に出てきた法制度の観点とか、人文・社会系の分析を人からITやAIに広げていったときの課題などに関するコメントとか、情報系と人文・社会系との連携・総合知に向けた今後の議論を深めていくにはどういうやり方がよいのかとか、幅広くいずれかに関して言いたいことやアドバイスがあれば順番に聞きたい。

大屋：神嶌さんから出た話に関連していうと、私はトロッコ問題の話がAI関連で出てくる度に、あれは人間も解けていないという話を強調する。人間も解けていないものがAIに解けるわけがないだろうという話である。この教訓は二つある。一つはAIと人間の違いはどこにあるかということ。現時点では少なくとも責任の取り方に違いがある。つまり、AIを拷問して痛めつけることはできないが、人間は監獄に閉じ込めてひどい目に遭わせることができるということである。この違いに着目したほうが恐らく有益だと思う。もう一つは、性能の違いは結局のところトラストに反映しない、ということである。TrustworthinessとTrustを分けて議論しているのがそれを表しているわけで、客観的な性質としてどれだけTrustworthyであっても、人々が現実に信用しないことがある。例えば原発というのはその例かもしれない。福島第一原発の事故後、説得力がなくなってしまったが、それまでは非常に安全性が高いけれども、トラストはないという状態だった。一方、餅ほど危険なものはないと思うのだが、みんな食べている。このように、性能とトラストの間には関連はあるけれども、必然はないということを前提に分析しなければいけないので、人間心理とか社会制度とかの観点からの検討が不可欠だと思う。そして、これをフィードバックする形で、いわゆる理系側のトラスト研究を進展させることが必要だと思っている。

村山：発表を聞いて、二つの観点を自分で考えようと思った。山口先生の発表にもあったように、デジタルトランスフォーメーションで新たな犯罪者が生まれたということが一つ目である。今までも、だます人はたくさんいたわけで、今までの犯罪者がデジタルトランスフォーメーションによって形を変えて出てきた。一方、犯罪学は専門ではないが、そういうことをする人が変わってきたのかなとも思う。それによって、脅威が少し前と違ってきたのではないかと思う。実際の物理的な犯罪のワールドとサイバースペースでの犯罪がちょっと違っていると感じた。もう一つは小山先生の話にあったように、トラストはすごく幅が広くて、いろんなものがあるということである。それをデジタルワールドに持ってくるには、スコープ・オブ・トラストと私は呼びたいのだが、「ここではトラストのこの観点について考えよう。」といった、もう少し狭めた定義があってもいいのではないと感じた。

　それから、最後、山本先生の発表は個人的に興味深い。なぜかというと、今回の話題ではないのだが、今、私は障害学を勉強し始めているからである。障害者とそれを支えるITの人たちの間のコミュニティーはあまりないので、PPI（Patient and Public Involvement）について非常に興味を持って聞いた。

山本：人間関係の信頼であるカスタマイズトラストと、社会を信頼するジェネラライズトラストの区別が大事ではないかと思う。フェイクニュースの話だが、この間の学会で英国での話を聞いた。コロナウイルスのワクチンを怖くて受けたくない人たちを調査したところ、彼らはフェイクニュースを信じていた。ただし、そういうフェイクニュースは皆が見ているわけで、どうしてあるグループはそれを信じて、別のグループは無視するのかということを考えると、学生の場合は自分が所属しているグループのトラストとの関係も大事なのではないかと思う。医療関係のフェイクニュースに関して、フェイクニュースを信じる人

の特徴や、その特徴はいつもそうなのか、いくつかの条件がそろうとだまされやすくなるのか、技術全体に信頼がないわけではないだろうが、いろいろな技術を使っているのでそこに問題があるのか、どうして特にこういった問題で普通の社会の感覚とは離れていくのか、といったことに関してもっと深い研究が必要ではないか、次のステップが必要ではないかと思っている。

小山：中川先生のコメントにあった、真正性とは何が真正かと言い出したらきりがなくて、定義が難しいという、真正性のことが以前から気になっている。トラストに関しては何か真正なものがあるというような理解をすべきではなく、真正性を追求しているけれども、真正なことがはっきりしていなくてもオーケーである、というのがトラストではないかと思う。

以前のワークショップで村山先生が言っていたが、ディストラストとか不信とか、信用できないという状況で信頼が崩れているということは分かる。つまり、こういうものは真正ではないということが分かる場面をうまく回避する、真正ではないということが判明したときに、そこをどうにか処理するためのものがトラストであって、真正なものがあって、それ以外のものは全て駄目だとするような構図でトラストを捉えるべきではない、という気がする。

そういう意味で、真正性という言葉は趣旨としてはそれを目指すものだけれども、定義せずに、まずありきなものではない、というように理解しないと、まずいのではないか。特にトラストというものが本来、分からないものの中でリスクを受け入れるというものであるという、その要素を考えると、とても重要なポイントになるのではないかと思った。

福島：冒頭の私のプレゼンテーションでも話したが、今日はトラストの三つの側面という話と、情報系から人文・社会系まで非常に幅広い取り組みと、アプリケーションという面でもいろいろな取り組みがあるという話まで、非常に幅広い話題を取り上げて開催した。トラストというもの自体が非常に幅広い側面を持ち、だから総合知的に取り組まなければならないという感覚を伝えたかったことから、まずはこのような形でのキックオフとした。今後、トラストについてシリーズ的にワークショップやセミナーのようなものを継続してやっていくことで、分野横断的な議論ができる場を作っていきたいと思っている。そのときは、例えばフェイクニュースの問題だったり、ブラックボックスAIの問題だったり、医療の問題だったり、あるいは対象真正性を中心にどう広げていくかという問題だったりとか、いろいろな切り口が考えられる中から、各回はそのどれかに絞って深い議論ができるように企画したいと思っている。トピックや問題点をフォーカスした議論を、技術開発に取り組んでいる人たちだけではなくて、人文・社会系の人たちも含めて広い観点からの議論ができるような場を設定していきたいと思っている。企画できたらご案内するので、興味がある方はぜひ参加していただきたい。

付録1　開催概要

日程：2023年1月10日（火）13：30～18：00
場所：Zoom Webinarによるオンライン開催

プログラム：　（敬称略）

第1部［30分］開催趣旨		
13：30～13：35	開催挨拶	木村康則（JST CRDS）
13：35～14：00	講演1：戦略プロポーザルの紹介	福島俊一（JST CRDS）
第2部［90分］トラスト3側面への取り組み事例		
14：00～14：30	講演2（対象真正性）Trusted Web	クロサカタツヤ（株式会社 企）
14：30～15：00	講演3（内容真実性）メディアのトラスト	山口真一（国際大学GLOCOM）
15：00～15：30	講演4（振る舞い予想・対応可能性）システムのトラスト	中島震（国立情報学研究所）
休憩［10分］		
第3部［80分］トラスト3側面の融合・分野横断の必要性		
15：40～16：00	講演5：トラスト3側面の捉え方	大屋雄裕（慶應義塾大学法学部）
16：00～16：20	講演6：情報系での国際的動向と分野横断の必要性	村山優子 （津田塾大学 数学・計算機科学研究所）
16：20～16：40	講演7：ステークホルダー関与、医療分野での実践事例	山本ベバリーアン （大阪大学 人間科学研究科）
16：40～17：00	講演8：人文・社会系での取り組みと分野横断の必要性	小山虎（山口大学 時間学研究所）
第4部［60分］総合討論		
17：00～18：00	司会	福島俊一（JST CRDS）
	総括コメンテーター	中川裕志 （理化学研究所 革新知能統合研究センター）
	質疑・議論	登壇者、聴講者
18：00	閉会	

参加登録者（登壇者と運営メンバーを除く）：150名

付録2 登壇者プロフィール

クロサカ タツヤ　株式会社 企（くわだて）代表取締役
1975年生まれ。慶應義塾大学・大学院（政策・メディア研究科）修了後、三菱総合研究所にて情報通信分野のコンサルティングや国内外の政策調査等に従事。その後2007年に独立し、現在は株式会社企（くわだて）代表として、通信・メディア産業の経営戦略立案や資本政策のアドバイザー業務を行う。2016年より慶應大学大学院政策・メディア研究科特任准教授。

山口 真一　国際大学GLOCOM 主幹研究員・准教授
1986年生まれ。博士（経済学・慶應義塾大学）。2020年より現職。専門は計量経済学、ネットメディア論、情報経済論等。主な著作に『ソーシャルメディア解体全書』（勁草書房）、『正義を振りかざす「極端な人」の正体』（光文社）等。KDDI Foundation Award貢献賞等を受賞。東京大学客員連携研究員、シエンプレ株式会社顧問、株式会社エコノミクスデザインシニアエコノミスト等も務める。

中島 震　国立情報学研究所 名誉教授
1981年東京大学大学院修士課程（物理学）修了。博士（学術）。NEC、法政大学を経て、2004年国立情報学研究所教授。2021年同名誉教授。ソフトウェア工学の研究に従事。現在、放送大学客員教授、産業技術総合研究所招聘研究員など。著書に『SPINモデル検査』（近代科学社、2008年）『ソフトウェア工学』（共編著、放送大学教育振興会、2019年）『AIリスク・マネジメント』（丸善出版、2022年）などがある。

大屋 雄裕　慶應義塾大学法学部 教授
1974年生まれ。慶應義塾大学法学部教授、専攻は法哲学。東京大学法学部を卒業、同大学助手・名古屋大学大学院法学研究科助教授・教授等を経て2015年10月より現職。著書に『自由とは何か：監視社会と「個人」の消滅』（ちくま新書、2007年）、『自由か、さもなくば幸福か？：21世紀の〈あり得べき社会〉を問う』（筑摩選書、2014年）、『法哲学』（共著、有斐閣、2014年）等がある。

村山 優子　津田塾大学 数学・計算機科学研究所 特任研究員
津田塾大学学芸学部数学科卒、三菱銀行および横河ヒューレット・パッカード社に勤務。1984年ロンドン大学内University College London大学院理学部計算機科学科修士課程修了。1992年同大学大学院博士課程修了（Ph.D.）。広島市立大学等を経て、1998年より岩手県立大学ソフトウェア情報学部助教授、2002年教授、2016年定年退職。2016年津田塾大学教授、2019年定年退職。2019年より現職。

山本 ベバリー アン　大阪大学 人間科学研究科 教授
英国ロンドン生まれ。英国国立シェフィールド大学大学院社会科学研究科Ph.D.修了。2006年より大阪大学人間科学研究科、2013年教授。2019年よりUniversity of Oxford, Centre for Health, Law and Emerging Technologies（HeLEX）, Academic Affiliate（兼任）。NPO法人HAEJ（遺伝性血管性浮腫患者会）理事長。JST RISTEXのAIとヘルスケアに関するAIDEプロジェクトの代表。

小山 虎　山口大学 時間学研究所 准教授

大阪大学大学院人間科学研究科博士課程修了。博士（人間科学）。大阪大学人間科学研究科・基礎工学研究科を経て、2017年より山口大学時間学研究所、2022年准教授。専門は分析哲学、形而上学、応用哲学、ロボット哲学。2018年に『信頼を考える：リヴァイアサンから人工知能まで』（勁草書房）編著。

中川 裕志　理化学研究所 革新知能統合研究センター チームリーダー

1975年東京大学工学部卒、1980年東京大学大学院工学系研究科修了（工学博士）。横浜国立大学工学部を経て、1999年より東京大学情報基盤センター教授、2018年より理化学研究所、現在に至る。東京大学名誉教授。研究テーマは、AI倫理、プライバシー保護の技術と法制度。著書に『裏側から視るAI－脅威・歴史・倫理』（近代科学社、2019年）『東京大学工学教程：機械学習』（丸善、2016年）等。

木村 康則　科学技術振興機構 研究開発戦略センター 上席フェロー

1981年富士通（株）入社。第五世代コンピュータ、京スーパーコンピュータなどコンピュータシステムの研究開発を経て、米国駐在。シリコンバレーにて研究マネジメントと研究成果の事業化に従事。2017年より現職。産業技術総合研究所招聘研究員兼任。その間、スタンフォード大学客員研究員、東京大学客員教授、九州大学客員教授等を歴任。オーム技術賞、情報処理学会業績賞を受賞。博士（工学）。

福島 俊一　科学技術振興機構 研究開発戦略センター フェロー

1982年東京大学理学部物理学科卒業、NEC入社。以来、中央研究所にて自然言語処理・サーチエンジン等の研究開発・事業化に従事。工学博士。2011～2013年東京大学大学院情報理工学研究科客員教授（兼任）。2016年より現職、人工知能分野を中心に研究動向調査・戦略提言を担当。情報処理学会論文賞、情報処理学会坂井記念特別賞、オーム技術賞等を受賞。情報処理学会フェロー。

作成メンバー

開催責任者	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）
企画・執筆とりまとめ	福島 俊一	フェロー	（CRDS システム・情報科学技術ユニット）
執筆メンバー	青木 孝	フェロー	（CRDS システム・情報科学技術ユニット）
	住田 朋久	フェロー	（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）
	花田 文子	フェロー	（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）

公開シンポジウム報告書 **CRDS-FY2022-SY-02**

デジタル社会における新たなトラスト形成

～総合知による取り組みへ～

令和 5 年 3 月　March 2023
ISBN 978-4-88890-831-3

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 978-4-88890-831-3

FOR THE FUTURE OF
SCIENCE AND
SOCIETY

CRDS

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