

2.4.7 社会におけるトラスト

（1）研究開発領域の定義

トラスト（信頼）を「相手が期待を裏切らないと思える状態」と定義する¹⁾。トラストは、情報科学分野だけでなく、心理学・社会学・政治学・哲学などの人文・社会科学分野も含め、幅広い分野で研究されており、その定義もさまざまだが^{2), 3)}、それらに共通したトラストの性質として次の4点が挙げられている⁴⁾。上記の定義はこれらの性質を踏まえたものである。

- ・トラストする側とトラストされる側という2者間の関係である。
- ・トラストは危険な状況や不確実な状況に存在する。
- ・トラストはリスクを取る行動につながる。
- ・トラストは非常に主観的な問題で、個人と環境の状況・文脈の影響を受ける。

トラストするか否かは最終的に各人の主観的な判断になるが、その判断に関わる「社会的よりどころ」を与え、人々がそれを活用して判断できるようにすることで、トラスト関係が社会に広がる。その際、セキュリティー・トラスト区分の他の研究開発領域で取り上げられているさまざまな要素技術が活用されるが、個々の技術内容は各研究開発領域にて記載されるので、本研究開発領域では、それらが社会に受容され、デジタル社会におけるトラスト形成の仕組みとして、うまく機能するようにするための研究開発を中心に記載する。

（2）キーワード

トラスト、信頼、安心、リスク、対象真正性、内容真実性、振る舞い予想・対応可能性、デジタル社会、制度設計、社会受容、総合知

（3）研究開発領域の概要

【本領域の意義】

冒頭で述べたように、トラスト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^{5), 6), 7), 8)}。

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てしまった。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。

本領域が目指すのは、デジタル社会におけるトラスト形成の仕組み作りによって、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である（図2-4-3）。

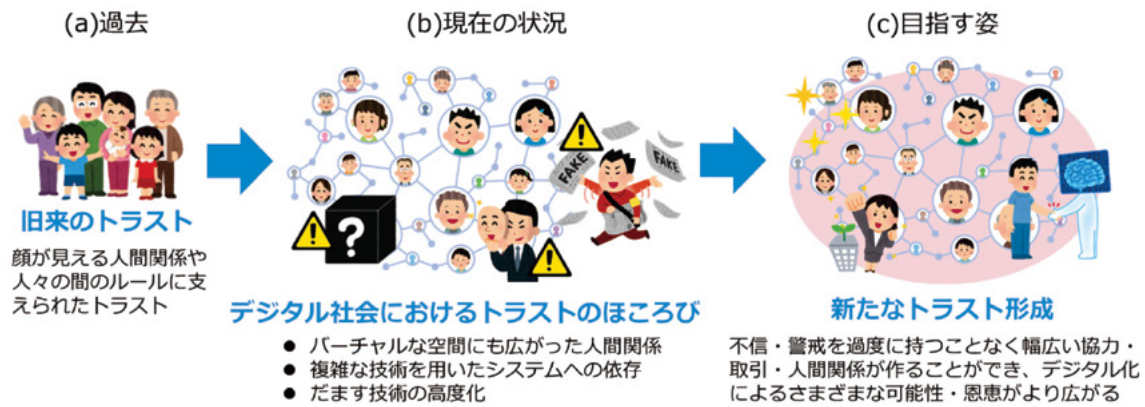


図2-4-3 デジタル社会におけるトラストの問題意識と目指す姿

〔研究開発の動向〕

① トラスト研究の系譜

トラストは社会秩序、人間関係、ビジネス上の取引など、さまざまな形で社会における重要な役割・効果を果たしていることから、トラストに関わる研究開発も、幅広い分野でさまざまな取り組みが進められてきた^{2), 3)}。

まず、人文・社会科学分野におけるトラスト研究として、古くは17・18世紀頃から哲学・社会学の分野で社会秩序問題という面から捉えた研究の流れがある。20世紀半ば頃には、心理学の立場から「囚人のジレンマ」問題を含む社会集団内の紛争解決要因としてトラストを位置付けた研究が進んだ。その後、1980年頃以降は、トラストがさまざまな役割の中で捉えられるようになってきた。Niklas Luhmann（ニクラス・ルーマン）がトラストの役割を「社会的複雑さの縮減」と位置付けたのは有名である。また、Robert Putnam（ロバート・パットナム）やFrancis Fukuyama（フランシス・フクヤマ）はトラストを社会関係資本（Social Capital）として位置付けている。Anthony Giddens（アンソニー・ギデنز）のリスク社会論や、近年のBruno Latour（ブリュノ・ラトゥール）らによるアクターネットワーク理論との関係も深い。国内では、社会心理学の立場から、認知バイアスとしての安心と信頼（トラスト）を対比して論じた山岸俊男の研究が知られている。人文・社会科学分野では、これまで人と人の間のトラストに関する研究が主だったが、近年は、情報ソースとトラスト、コンピューターエージェントとトラスト、トラストと非言語行動、トラストと社会認知など、社会心理学におけるトラスト研究の対象や尺度が拡大してきている。

次に、情報科学分野では、「トラスト」という言葉を使った研究開発が始まったのは1990年代以降である。情報科学技術分野でのトラストと言ったとき、セキュリティーやプライバシーの研究に関わる取り組みが一つ大きな流れになっている。デジタルトラスト、トラストサービスといった取り組みは、デジタルアイデンティティーや認証基盤などの仕組みをベースに個人・組織などが本人・本物であることを保証しようとするものである。また、その下位のコンピューティング層では、CPUなどのデバイスが正しく動作することを保証しようというDevice Integrityの取り組みがある。これらは、コンピューターとネットワークを用いたさまざまなデジタルサービスにおけるトラストを支える基盤技術となっている。そこに関わる新しいパラダイム、特に改ざんや盗聴を防止するための新技術として、近年、ブロックチェーン技術や量子暗号（量子鍵配送）技術も注目されている。一方、アプリケーション層では、電子商取引、災害時コミュニケーションなどにおいて、トラストの感情面を考慮した取り組みがある。また、AIやロボットと人間との間のトラスト関係が、近年注目されるトピックとなっている。

さらに、トラスト研究の一面として、科学技術のリスク面に目を向けた取り組みがある。ITシステムのリスクに関する考え方として、1980～1990年代はコンピューター安全という面からハードウェアや基盤ソフ

トウェアの信頼性が着目された。2000年代になると、それにアプリケーションまで含めた信頼性・安全性を考えるようになり、ソフトウェアディペンダビリティという見方がされるようになった。さらに近年は、AIやCPS（Cyber-Physical Systems）のTrustworthinessとして、信頼性・安全性に加えて、回復性・プライバシー・セキュリティなども併せて論じられるようになった。

一方、1975年のアシロマ会議に始まり、科学技術のELSI（Ethical, Legal and Social Issues:倫理的・法的・社会的課題）面の議論が活発に行われるようになり、近年、ITシステム関連では特にAI ELSIが重要課題になっている。「信頼されるAI」「Trustworthy AI」といった表現が用いられ、AI社会原則・AI倫理指針が国・国際レベルで掲げられるようになった。国際標準化活動においてもAIのTrustworthinessが取り上げられている。また、デマや偽情報は古くから存在したが、インターネットやソーシャルメディアの普及・発展に伴い、フェイクニュースが社会問題化し、ディープフェイクで生成されたフェイク動画などのAIによるフェイクの高度化に懸念も高まっている。

② 現在のトラスト研究の概観

上述のようなトラスト研究の系譜・広がりを経て、現在取り組まれている主要な研究開発トピックを5つの分野ごとに概観する。より具体的な例示を表2-4-1に示す。

表2-4-1 トラスト研究の分野と研究開発トピック例

分野	研究開発トピック例
A: デジタルトラスト	トラストアンカー/トラストチェーン、ブロックチェーン、認証局、タイムスタンプ、eシール、電子署名、生体認証、分散型アイデンティティ、Remote Attestation、Confidential Computing、Trusted Execution Environment（TEE）、Hardware Root of Trust、Trusted Boot/Secure Boot、Trusted Communicationなど
B: フェイク対策	フェイク検知、ファクトチェックなど
C: 信頼されるAI	機械学習品質マネジメントガイドライン、機械学習テスト手法、Assured Autonomy、説明可能AI（XAI）、Safe Learning、公平性配慮機械学習、プライバシー配慮機械学習など
D: AIガバナンス	AIガバナンス、アジャイルガバナンス、ガバナンスエコシステム、リスクチェーンモデルなど
E: トラストの観察・理解	トラストの非対称性、能力・意図モデル、ABIモデル、SVSモデル、主観的確率としての信頼、安心vs.信頼の理論、信頼尺度・信頼計測、社会関係資本とトラスト、協調行動の信頼・規範ネットワークなど

「A：デジタルトラスト」は、セキュリティ技術との関係が深く、現在特に活発に取り組まれている分野である⁹⁾(詳細は「2.4.5 システムのデジタルトラスト」「2.4.6 データ・コンテンツのデジタルトラスト」に記載)。ハードウェア的な改ざん防止の仕組み、ブロックチェーン（改ざん防止機能を持つ分散型台帳）、暗号技術に基づく電子署名・認証、個人や機器の認証技術などが含まれる。

「B：フェイク対策」には、近年社会問題化しているフェイクニュースやフェイク画像・動画・音声などの検知や拡散防止のための技術や、ファクトチェックの活動・システムなどが該当する^{10), 11)}(詳細は「2.1.5 人・AI協働と意思決定支援」に記載)。デジタルトラストと比べると、何が正しいかを定めることが難しく、「表現の自由」の問題に関わることも多い。

「C：信頼されるAI」(Trustworthy AI)も近年非常に活発に取り組まれるようになった分野で、AIシステムの安全性・信頼性・社会受容性を確保するためのさまざまな技術開発が進められている^{12), 13)}(詳細は「2.1.4 AIソフトウェア工学」に記載)。ブラックボックスとも言われる深層学習（Deep Learning）などのAI技術による判定結果に説明を与える説明可能AI（Explainable AI: XAI）技術、機械学習技術を用

いた帰納型のシステム開発におけるテスト・デバッグ手法や品質管理ガイドラインなどソフトウェア工学的な方法論、プライバシーや公平性を確保するデータ分析法など多岐にわたっている。

AIシステム開発のための技術分野とは別に、運用に関わるガイドラインや活用に関わるルールの整備などを含めた「D：AIガバナンス」と呼ばれる取り組みも進められている¹⁴⁾（詳細は「2.1.9 社会におけるAI」に記載）。

また、哲学・社会学・政治学・心理学・経済学などの分野では、トラストの定義・モデル化、人間の振る舞い・態度に関する観察・比較実験などが行われてきた^{3), 5), 15)}。この「E：トラストの観察・理解」の研究では、長年、人間に対するトラストが論じられてきたが、近年は機械・システムに対するトラストも検討対象に含められるようになってきた。

以上のように、トラストに関連する異なる切り口からさまざまな研究開発が進められている。研究論文数が近年大きく増加している研究開発トピックも多い。しかし、さまざまな分野で進められているトラスト研究開発の間での知見共有・連携はほとんど見られない。多くの研究開発は、トラストに関わるある一面にフォーカスしたものであることから、個別的な対処や断片的な状況改善にとどまりがちである。この状況は、国内だけでなく海外でも同様である。本節の冒頭で述べたようなデジタル社会におけるトラスト形成の仕組み作りに向けては、上述のA～Eのような個々の研究開発を一層推進するとともに、それらを総合した社会的トラスト形成への取り組みが望まれる¹⁾。

③ 関連政策・プログラムの状況

上記②に示した分野ごとに関連政策・プログラムとして注目されるものを表2-4-2に示す。
これらのうち特に注目される取り組みとして、内閣官房デジタル市場競争本部の「Trusted Web 推進協議会」と、英国研究・イノベーション機構（UKRI）の「Trustworthy Autonomous Systems Programme」を〔注目すべき国内外のプロジェクト〕の項で取り上げる。

表2-4-2 トラスト関連政策提言・プログラム事例

分野	国	政策提言・プログラム事例
A：デジタル トラスト	日本	・「Data Free Flow with Trust (DFFT)」(2019年1月ダボス会議、安倍元首相) ・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」(2022年8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～2022年7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	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	日本	・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」、さががけ「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	・英国研究・イノベーション機構(UKRI)「Trustworthy Autonomous Systems Programme」

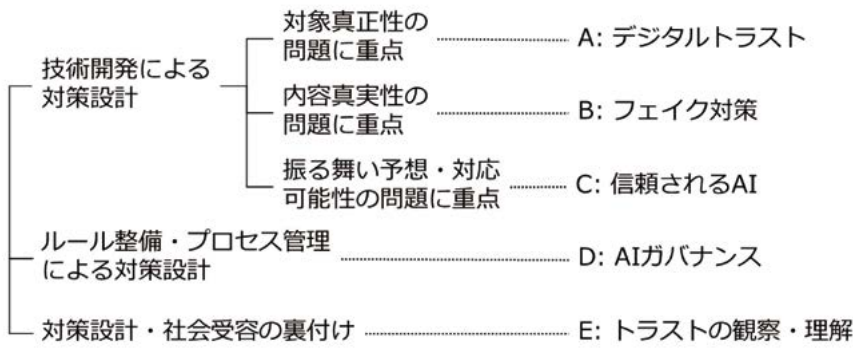
D: AIガバナンス	日本	<ul style="list-style-type: none">・世界経済フォーラム「Rebuilding Trust and Governance: Towards DFFT」白書 (2021年3月)・経済産業省「Governance Innovation Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」(2021年7月)、「AI原則実践のためのガバナンス・ガイドライン Ver. 1.1」(2022年1月)・日本ディープラーニング協会「AIガバナンスとその評価」研究会報告書「AIガバナンス・エコシステム-産業構造を考慮に入れたAIの信頼性確保に向けて-」(2021年7月)
	欧州	<ul style="list-style-type: none">・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)

(4) 注目動向

[新展開・技術トピックス]

① トラストの3側面と多面的・複合的な検証

[研究開発の動向] で取り上げたさまざまな研究開発では、対象（トラストする相手）のどのような側面を扱っているかに違いが見られる。これを整理するため「トラストの3側面」という概念が示された¹⁾。すなわち、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?）を「トラストの3側面」と呼ぶ。また、トラスト問題への対策は、技術開発による対策だけでなく、制度設計（ルール整備・プロセス管理などを含む）による対策も考えられ、また、それらが有効に働くかの裏付けも求められる。以上を踏まえて、トラスト研究開発に関する前述の5分野を整理すると、図2-4-4のようになる。[研究開発の動向] の項で「さまざまな分野で進められているトラスト研究開発の間での知見共有・連携はほとんど見られない」と述べたが、このような観点の違いが一つの要因と考えられる。



[注] おおまかな傾向であり、この見方に収まらない取り組みも存在する

図2-4-4 トラスト研究の観点の違い

しかし、人が主観的に判断する場合には、おそらく3側面を多面的・複合的に捉えた上で、トラストできるかを総合的に判断していると思われる。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、「そのサービスの仕組み（どのように動いてどのような結果が得られそうか）が信じられるか」（振る舞い予想・対応可能性の側面）、「そのサービスの提供企業が怪しくないか」（対象真正性の側面）、「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」（内容真実性の側面）というように、多面的にチェックしようとするであろうし、一つの側面についても複数の情報を突き合わせることもするであろう。このように、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というようなことを、人は行っている。

「トラストの3側面」という整理に基づき、今後、トラストに関するより総合的な研究開発や戦略構築が進むと期待される。

② Web3

ハイパーテキスト構造をベースとした読むことが主体の第一世代 Web (Web 1.0)、ソーシャルネットワークサービス (SNS) など双方向のメディアとなった第二世代 Web (Web 2.0) に対して、ブロックチェーンを用いた分散型の Web3 が次世代 Web として注目されている。Web3 は Web 3.0 と称されることも多いが、セマンティック Web の概念を中核として提唱された Web 3.0 とは異なるものである。2022 年 9 月、デジタル庁は「Web 3.0 研究会」(座長：慶応義塾大学の國領二郎教授) を設置し、2022 年 12 月に「Web 3.0 研究会報告書～Web 3.0 の健全な発展に向けて～」(Web3.0 研究会) を公表している。

Web3 自体は、必ずしもトラストを打ち出した概念ではないが、後述する Trusted Web と一部類似するところがあり、今後のデジタル社会基盤に関連した動向として注目される。「Trusted Web ホワイトペーパー Ver.2」⁸⁾ では、「昨今、次世代のインターネットやウェブのあり方として、「Web3」という概念が広く議論されている。現状のインターネットやウェブに対する問題意識や、分散型で検証可能な部分を広げること志向しているという意味での方向性は、Trusted Web と共通するものがあると考えられるが、「Web3」の厳密な定義についてはさまざまな見解があり、定義は定まっていないと考えられる。こうした中で、Trusted Web については、アイデンティティ管理のあり方に重点を置くほか、技術中立的な取り組みとして進めているものであり、ブロックチェーン技術の活用のみでなく、検証可能性を高めるさまざまな枠組みを活用し、組み合わせることにより、Trust のレベルを高めることを目指すものである。」との言及がある。

[注目すべき国内外のプロジェクト]

① Trusted Web 推進協議会

「Data Free Flow with Trust (DFFT)」(2019 年 1 月ダボス会議) や「デジタル市場競争に係る中期展望レポート」(2020 年 6 月デジタル市場競争会議) を受けて、2020 年 10 月に内閣官房デジタル市場競争本部に Trusted Web 推進協議会 (座長：慶応義塾大学の村井純教授) が設置された。

そのコンセプトや設計方針などは「Trusted Web ホワイトペーパー」(2021 年 3 月に Ver.1、2022 年 8 月に Ver.2 を公開)⁸⁾ にまとめられている。Trusted Web では「やり取りされるデータが信頼できるか、データをやり取りする相手方を信頼できるか、提供したデータの相手方における取り扱いを信頼できるか」といった点をペインポイントと捉えて、インターネットや Web の良さを活かし、その上に重ね合わせるオーバーレイアプローチを採る。特定サービスに過度に依存せず、ユーザー (自然人または法人) 自身が自らに関連するデータをコントロールすることを可能とし、データのやり取りにおける合意形成の仕組みを取り入れ、その合意の履行のトレースを可能としつつ、検証できるケースを拡大することによって、トラストの向上を目指している。

ホワイトペーパー Ver.2 ではユースケースも取り上げられたが、より具体化・検証を進めるべく、「Trusted Web の実現に向けたユースケース実証事業」が公募され、2022 年 9 月に 13 案件が採択された。

② Trustworthy Autonomous Systems Programme (TAS プログラム)

TAS プログラムは、英国研究・イノベーション機構 (UK Research and Innovation: UKRI) による「トラストできる自律システム」に関する学際的研究プログラムで、ELSI (Ethical, Legal and Social Issues: 倫理的・法的・社会的課題) や RRI (Responsible Research and Innovation: 責任ある研究・イノベーション) の諸課題を包括している。資金規模は約 52 億円で、その体制は、Verifiability (検証可能性)、Governance and Regulation (ガバナンスと規制)、Trust (信頼)、Security (安全保障・人権

なども含むセキュリティー)、Resilience (レジリエンス)、Functionality (機能性) という6つのノードとハブで構成され、主要大学と多数の産業界からの参加がある。自動運転、医療・介護、防衛・安全保障、AI 倫理とガバナンスなどにわたり、12 件のプロジェクトが採択されている¹⁶⁾。前述の分類C (信頼されるAI) に重点を置いたプログラムだが、大型の学際的研究体制を構築している点が注目に値する。

(5) 科学技術的課題

本領域が目指す、デジタル社会におけるトラスト形成の仕組み作りに向けて、以下の4 層から成る研究開発課題が重要と考えられる。これらは相互に関連し、連携した取り組みが有効である。

① トラストの社会的よりどころの再構築

表2-4-3に、「トラストの3側面」のおのおのについて、現状の「社会的よりどころ」としてどのようなものが機能しているかを例示するとともに、デジタル化の進展に伴い、それらでは不十分になってきているという問題点も併せて示した。例えば内容真実性について、従来は証拠写真・監視カメラ映像などが事実性判断の「社会的よりどころ」になり得たが、AI 技術によって高品質なフェイク映像が簡単に生成できるようになってしまったため、必ずしも確かなよりどころにならなくなってしまったというのが一例である。

このような問題に対して、トラストの3側面 (対象真正性/内容真実性/振る舞い予想・対応可能性) で何を社会的よりどころに設定するか、社会的よりどころをどのような技術と制度によって担保するか、を考えねばならない。対象真正性/内容真実性/振る舞い予想・対応可能性それぞれの社会的よりどころの再構築、複合的検証のメカニズム、改ざんされない記録・トレーサビリティなどの研究開発が含まれる。

表 2-4-3 トラストの3側面に対する「社会的よりどころ」の例と問題点

トラストの3側面	現状の社会的よりどころ	問題点
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。トラスト基点の信頼性担保にも課題あり。
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。そもそも絶対的の真実・事実は定まらず、ファクトチェック可能な対象は限定的。
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。

② 社会的トラスト形成フレームワーク

①の「社会的よりどころ」を用意するだけでなく、人々がそれを容易に使いこなし、トラストできる対象を広げていけるようにするとともに、「社会的よりどころ」が公正・健全に維持されるようにすることも、重要な研究開発課題である。トラスト域拡大と権限制御、公正・健全なトラスト基点の維持、トラストの悪用・攻撃への対策、使いこなしを容易にする技術・教育などの研究開発が含まれる。

③ 具体的トラスト問題ケースへの取り組み

デジタル化の進展で生じた具体的なトラスト問題の各ケースに対して、上記①②の枠組みを用いた解決や状況改善を実証する取り組みも重要である。具体的ケース固有の問題分析・対処と具体的ケースからの①②の研究開発へのフィードバックも含む。トラスト問題ケースの例としては、ネット取引・サプライチェー

ンやメタバースなどに関わる「ビジネスにおけるトラスト」、フェイクニュースやインフォデミックなどの問題に関わる「ネット情報のトラスト」、自動運転車やAIエージェントなど「AI応用システムのトラスト」、医療者と患者との関係にAIがセカンドオピニオン的に関わるなどの「専門家+AIのトラスト」などが挙げられる。

④ トラストに関する基礎研究

上記①②③の実現とその社会受容のため、社会におけるトラストについての理解や、そのデジタル化による影響・変化に関する基礎的な研究開発も重要である。例えば、デジタル社会におけるトラスト形成や不信のメカニズム理解、トラストに関わる日本人のメンタリティーと国際比較・文化差、デジタル社会のトラスト形成のための方策・対策設計の裏付けなどに関する基礎研究が挙げられる。

(6) その他の課題

① 分野間の知見共有・連携促進の場作り

トラスト研究には「総合知」¹による取り組みが不可欠であり、分野横断の学際的研究の活性化が望まれる。トラストに関わる研究分野の幅広さに対して、研究者個人の人脈や出会いを通してボトムアップにリーチできる範囲には限界があることから、分野間の知見共有・連携促進の場作り²⁾を、トップダウンな施策によって立ち上げることが有効であろう。このような場が、総合的なトラスト研究のビジョン共有、幅広い研究事例の把握、分野横断の共同研究のタネの発見機会となり、分野横断・学際的な研究コミュニティの活性化につながる。

② 学際的トラスト研究の継続的活動体制と人材育成策

学際的研究を骨太化し、研究者層を厚くしていくためには、上記①のような仕掛けをトリガーとしつつも、学際的トラスト研究の継続的な活動体制を作っていくことが必要である。研究活動の母体ができることで、研究人材の育成にもつながる。具体的には、学会・研究会の立ち上げ、このような学際的活動をコーディネートする人材やチームへの助成、学際的トラスト基礎研究を推進する拠点あるいはバーチャルな連携体制の構築などが、打ち手の候補として考えられる。

人材育成に関しては、分野横断・文理融合的な「総合知」に取り組むことに対する阻害要因が指摘されている。阻害要因の一例として、テニユアポストが少ないこと、「総合知」への取り組みは研究論文になりにくい一方で任期付きポストでは論文業績が必要になること、個別分野で業績を上げながら「総合知」に取り組まなければならないという高いハードルが現実といったことが挙げられる。これはトラスト研究に限った問題ではなく、根本的な打ち手は必ずしも容易なことではないが、考えていく必要がある。

③ ファundingプログラムによる推進・加速

上記②によってトラスト共通基礎研究を育成する一方で、より具体的な問題解決型の目標を設定し、それをファundingプログラムによって推進・加速することも必要であろう。前述の「トラストの3側面」のそれぞれに対する取り組みを表2-4-2に例示したが、さらにそのスコープを拡大し、総合的なトラスト形成に向けた連携を促進したり、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、メタバースなど具体的シーンでのトラスト問題に対策したりといった展開・発展が期待される。

日本政府はDFFT（Data Free Flow with Trust）を掲げたデータ戦略や、Trusted Quality AIを掲

1 「第6期科学技術・イノベーション基本計画」で掲げられた「多様な分野の知見を合わせて新しい知を創出する」という考え方。詳細は、「総合知」ポータルサイト（内閣府）を参照いただきたい。
<https://www8.cao.go.jp/cstp/sogochi/index.html>

げたAI戦略を発信している。上記①②③によって、これをさらに推し進めた、より総合的で一貫性のあるトラストを軸とした日本の研究戦略を構築でき、国際的にも先行したコンセプト発信が狙い得る。

参考文献

- 1) 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「戦略プロポーザル：デジタル社会における新たなトラスト形成」<https://www.jst.go.jp/crds/pdf/2022/SP/CRDS-FY2022-SP-03.pdf>, (2023年2月25日アクセス)。
- 2) 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『俯瞰セミナー&ワークショップ報告書 トラスト研究の潮流：人文・社会科学から人工知能、医療まで』(2022), <https://www.jst.go.jp/crds/pdf/2021/WR/CRDS-FY2021-WR-05.pdf>, (2023年2月25日アクセス)。
- 3) 小山虎 編著『信頼を考える：リヴァイアサンから人工知能まで』(東京：勁草書房, 2018)。
- 4) Piotr Pietrzak and Josu Takala, “Digital trust - a systematic literature review,” *Forum Scientiae Oeconomia* 9, no. 3 (2021) : 56-71., https://doi.org/10.23762/FSO_VOL9_NO3_4.
- 5) ニクラス・ルーマン『信頼：社会的な複雑性の縮減メカニズム』大庭健, 正村俊之 訳 (東京：勁草書房, 1990)。
- 6) レイチェル・ボッツマン『TRUST:世界最先端の企業はいかに〈信頼〉を攻略したか』関美和 訳 (東京：日経BP, 2018)。
- 7) World Economic Forum, “White Paper: Rebuilding Trust and Governance: Towards Data Free Flow with Trust (DFFT),” <https://www.weforum.org/whitepapers/rebuilding-trust-and-governance-towards-data-free-flow-with-trust-dfft/>, (2023年2月25日アクセス)。
- 8) Trusted Web 推進協議会「Trusted Web ホワイトペーパー Ver.2.0 (2022年8月15日)」首相官邸, https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/trustedweb.pdf, (2023年2月25日アクセス)。
Trusted Web 推進協議会「Trusted Web ホワイトペーパー Ver.1.0 (2021年3月12日)」首相官邸, https://www.kantei.go.jp/jp/singi/digitalmarket/trusted_web/pdf/documents_210331-2.pdf, (2023年2月25日アクセス)。
- 9) 日本ネットワークセキュリティ協会 PKI 相互運用技術 WG, 電子署名 WG「デジタル社会におけるトラスト」PKI & TRUST Days online 2021 (2021年4月15-16日), <https://www.jnsa.org/seminar/pki-day/2021/index.html>, (2023年2月25日アクセス)。
- 10) 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『公開ワークショップ報告書 意思決定のための情報科学：情報氾濫・フェイク・分断に立ち向かうことは可能か』(2020), <https://www.jst.go.jp/crds/pdf/2019/WR/CRDS-FY2019-WR-02.pdf>, (2023年2月25日アクセス)。
- 11) 山口真一『ソーシャルメディア解体全書：フェイクニュース・ネット炎上・情報の偏り』(東京：勁草書房, 2022)。
- 12) 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「戦略プロポーザル：AI応用システムの安全性・信頼性を確保する新世代ソフトウェア工学の確立」<https://www.jst.go.jp/crds/pdf/2018/SP/CRDS-FY2018-SP-03.pdf>, (2023年2月25日アクセス)。
- 13) 柿沼太一, 他『機械学習工学』石川冬樹, 丸山宏 編著, 機械学習プロフェッショナルシリーズ (東京：講談社サイエンティフィク, 2022)。
- 14) AI原則の実践の在り方に関する検討会, AIガバナンス・ガイドラインWG「AI原則実践のためのガバナンス・ガイドライン Ver. 1.1 (令和4年1月28日)」経済産業省, https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_1.pdf, (2023年2月25日アクセス)。
- 15) 山岸俊男『信頼の構造：こころと社会の進化ゲーム』(東京：東京大学出版会, 1998)。

16) UKRI Trustworthy Autonomous Systems Hub, “Annual Report 2021,” <https://www.tas.ac.uk/wp-content/uploads/2021/09/Annual-Report-2021.pdf>, (2023年2月25日アクセス) .

2.4

俯瞰区分と研究開発領域
セキュリティ・トラスト