第13回

トラスト研究の現場から
――情報科学の視点：研究開発戦略センターによる学際研究をもとに

はじめに

株式会社をはじめとするさまざまな組織体のパーパス（存在意義）として「信頼を構築する」（Building trust）、または単に「信頼」「トラスト」という表現や単語をしばしば見かけます。ここで使われている「トラスト」とは何を意味している概念なのでしょうか。

「トラストとは信頼である」などと訳してみて、それで納得してしまうこともありそうですが、本稿ではより進んでトラストについて考えてみたい方にヒントとなりそうなものを紹介します。トラストは古くは17世紀の哲学者ホッブズの時代から、永年にわたり数多くの「知の巨人」たちがチャレンジしてきた研究テーマであり、そこには広く深い世界が広がっています。

本稿では、国立研究開発法人科学技術振興機構（Japan Science and Technology Agency：JST）に属する研究開発戦略センター（Center for Research and Development Strategy：CRDS）が、数年前から取り組んでいる学際的研究「デジタル社会における新たなトラスト形成」の研究成果の中から、筆者が現代におけるトラストのエッセンスと考える部分について解説します。

なお、本稿における意見にわたる部分については筆者個人のものであり、所属するPwCあらた有限責任監査法人の公式見解ではないことを申し添えます。

1 トラスト（信頼）の定義および役割

トラストはさまざまな学問分野において研究がなされており、定義ひとつとっても学問領域、また研究者によって異なります。トラストをどう定義するのか自体が大きな研究テーマです。

ここでは多種多様な定義を羅列することは避け、トラストとは、おおむね「相手が期待を裏切らないと思える状態」と考えることにします。すなわち、「リスクはあるけれども信頼することにより安心して迅速な行動または意思決定が可能となること」です。これは、取引や協力のコスト削減につながるため、ビジネスにとっては非常に重要なことです。

そのうえで、トラストの機能を検討するためにはいくつかの学問領域における代表的な定義から、関連する他の用語との関係を考えることが役立つと考えられます。以下はその例です。

(1) 行動経済学・実験経済学

トラストと似た概念に「安心（Assurance）」があります。人々が安心する気持ちを持つ背景には、仮にその相手の人が自分を裏切って何かをしたとすると、相手自身が損をしてしまうという状況があります。その損得勘定ゆえに人々は安心することができるというわけです。

これに対してトラストの場合、その相手についていわゆる「いい人」であるとの感情を抱いているということ、つまり相手の人間性に焦点が当てられていることが多いと言えます。また、相手が自分に対しても好意を持っていると思えるということもあります。つま

り損得ではなく、相手との関係性を指すことがしばしばです。

まとめると、「安心が経済的な動機によるものである」のに対し、「信頼はより相手の人格に左右されるものである」との違いがあります。

(2) 情報科学

情報科学の領域ではトラストと類似または関連した概念として、「セキュリティ」「セーフティ」「信頼性」があります。以下にそれぞれの概念の意味、また相違点についてまとめています（**図表1**）。

図表1：情報科学におけるトラスト概念とその類似概念

概念	対象	相違点
セキュリティ（Security）	心的な脅威	故意による
セーフティ（Safety）	身体的な脅威	故意でない
信頼性（Reliability）	物理的な脅威	故意でない
参考：トラスト（Trust）	心的な脅威	故意による場合、および故意でない場合

出所：参考文献をもとにPwC作成

また、トラストの要素のそれぞれには客観的に測れる部分と感情的な部分があり、それらのうち感情的な部分、すなわち依存を意味するものとしてまとめて「安心」と呼ぶことがあると考えられます。

(3) 法律学

法律学でトラストというと、重要なのは「信託」の概念です。すなわち、①自己の財産を、②第三者に委託し、③自己または他者のために運用・管理する制度のことを意味します。

ここで、委託者と受託者の関係によって、信頼性を支えるものを「代理」「権威」「信託」の3つに分類できます。

特に委託者よりも受託者の知識や能力が高く（代理はその逆）、内容が定まっていない場合であって、かつ、あらかじめ国家によるライセンスなどによって統制（＝権威）することができないものについて、信任関係として現代における信託として扱えるのではないかと考えられます。

(4)（学問としての）人工知能

人工知能（AI）の分野では、トラストとの関係で「説明可能性」と言われますが、この言葉が独り歩きするのは好ましくなく、理解できるものであることが求められます。他方、透明性では補償する責任者、何か不都合なことが起こったときに誰が補償するかが求められます。アカウンタビリティとは、理解可能性と透明性の2つの概念が合わさって得られるものですが、ただ説明すればよいという日本の説明責任とは異なることに注意が必要です。

理解可能性については、AIの内部の動きを人間が理解できる形で説明することはほぼ無理であるため、近似的なモデルを作ることが行われています。分かりやすさを目指すものです。

他方、トラストについては、例えば「大会社だから」「みんなが使っているから」「自分と似たようなケースで同じ結果が出ているようだから」といった文脈で使われることが多いと考えられます。

2 デジタル社会におけるトラスト問題

旧来の古典的なトラストは、身近な人たちの信頼関係を中心に議論されてきました。しかし、デジタル化の進展により、バーチャルな人間関係の広がり、複雑な技術を用いたシステムへの依存、詐称などの手口の高度化などにより、トラストをめぐる環境変化が起きています。

そのような変化によって、トラストに関わる問題が生じつつあります。デジタル化の進展が生んだトラスト問題としてしばしば挙げられる具体例を以下の8つに類型化しました。

① 仮想世界のトラストに基づく取引

仮想世界や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が起きています。対策がとられているものの、常に新しい仕組みが登場し、新たなリスクが発生して対策が追いつかないケースが見受けられます。

② メディアにおけるフェイク拡散

フェイク動画による政治への干渉や個人攻撃等が社会問題となっています。簡単に人をだますことができ、裁判などでの証拠の信頼性も揺らぎます。他方、フェイクに対する法規制が強くなると、表現の自由が妨げられる恐れも生じます。

③ 自動運転車

AIの内部的な動きはブラックボックスで、動作保証や精度保証ができません。このため、安心して乗車できるか、事故発生時に原因究明や責任の所在はどうなるかが問われることとなります。

④ パーソナルAIエージェント

個人情報の管理代行の中身はブラックボックスであり、個人の意図や期待通りに振る舞うことを保証できないことがあります。そのため、個人情報を委ねることができるのか、期待に反する事態が起きた場合の責任所在がどうなるのかといった問題があります。

⑤ 人を評価するAIシステム

AIシステムの学習データに偏りや差別的な要因が含まれると、不公平で差別的な評価を助長します。また、評価アルゴリズムに過剰適合して行動する人々が生み出される恐れがあります。

⑥ 医療意思決定におけるAIセカンドオピニオン

異なる多様な医療情報を患者が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ります。三者間における新たな役割関係とそこでのトラスト関係の在り方が検討されなければなりません。

⑦ コミュニケーションロボット

製造技術の発達により、人間らしい外観を備えるようになったロボットに人間並みの能力を期待および過信してしまい、期待外れだったと失望する人が増えやすくなります。逆に、身近なロボットに過度な親近感や依存感を持ってしまうタイプの人もいます。

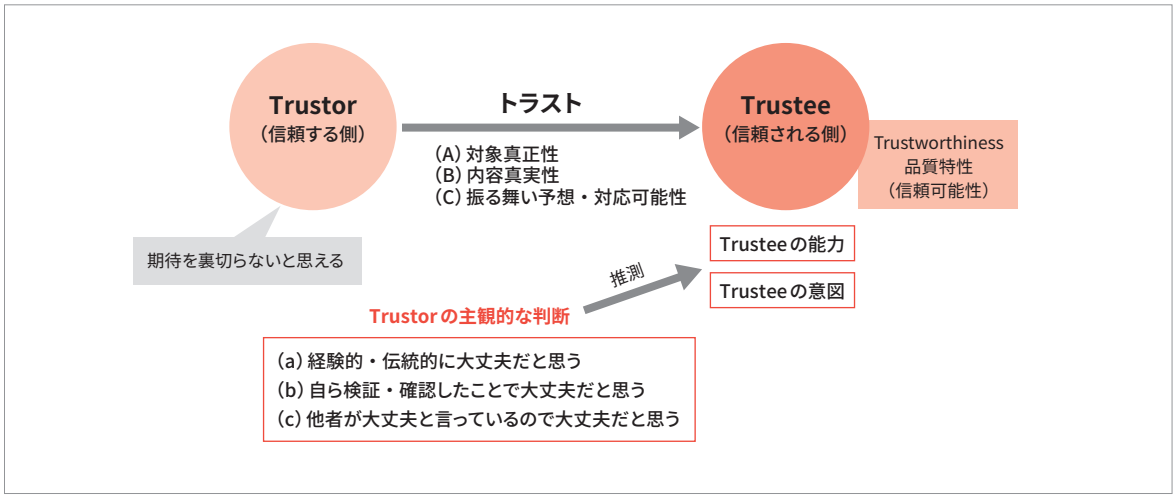
⑧ メタバース内活動におけるトラスト

生身の人間や物理的な実態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立するのか否かが問題です。

3 トラストのモデル

トラストのモデルの一例として、今回ご紹介している研究開発戦略センターのモデルを示します（**図表2**）。

図表2：トラストモデル



出所：参考文献をもとにPwC作成

まず、トラストは、「Trustor（信頼する側）」と「Trustee（信頼される側）」の関係であり、Trustorにとって、Trusteeは期待を裏切らないと思えるときに、トラストが成り立ちます。期待を裏切らないと思えるか否かは、Trustorの主観的な判断によることとなります。

その判断の際に着目され得るTrusteeの品質特性を「Trustworthiness（信頼可能性）」と呼びます。Trustworthinessを計測／観測／検証した結果から得た裏付けから「期待を裏切られない」と思えるケースもあれば、必ずしもそのような裏付けはなく、「リスクがあるが、大丈夫だ」とみなすケースもあります。

この2つのケースの比がどの程度ならば大丈夫と思えるかは、人によって異なります。

期待を裏切られない（大丈夫だ）と思えるかどうかを主観的に判断する際には、Trusteeの能力（期待に応える能力があるか）や意図（期待に応えようとしているか／悪意を持っていないか）などを推測して判断していると言われます。また、この主観的な判断にはいくつかのパターンがあります。

【Trustorの主観的な判断】
(a) 経験的・伝統的に大丈夫だと思う。
(b) 自ら検証・確認したことで大丈夫だと思う。
(c) 他者が大丈夫と言っているので大丈夫だと思う。

また、トラストには3つの側面があるとされます（**図表3**）。

図表3：トラストの3側面

側面	説明
(A) 対象真正性	本人・本物であるか？
(B) 内容真実性	内容が事実・真実であるか？
(C) 振る舞い予想・対応可能性	相手の振る舞いに対して想定・対応できるか？

ここで提示している「トラストは相手が期待を裏切らないと思える状態」という定義では、(C) の側面（相手の振る舞い）を中心に置いています。もっとも、そもそもその相手が本人・本物だという（A）の側面は大前提です。また、相手が人や組織やシステムという場合は、その振る舞いが期待を裏切らないかを考えますが、相手が情報という場合もあり、その場合は、その内容が事実・真実だという（B）の側面を期待します。

これは別な言い方をすると、相手が情報の場合、それを使って行動・意思決定をするのは自分であり、それによって自分が招いた結果が期待通りのものかどうかは、内容真実性に左右されるということです。

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・複合的に捉えたうえで、トラストできるかを総合的に判断していると思われます。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しょうし、1つの側面について複数の情報を突き合わせることもするでしょう。

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1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ことを、人は行っています。

4 TrustorとTrusteeに関わる要因とその変化

ここでは、TrustorとTrusteeのそれぞれに関係する要因と変化を4点挙げます。

(1) Trusteeの多様化

Trusteeとなるものには、個人だけでなく、集団、組織、政府、専門家コミュニティのような人が集まって形作られるものも該当しますし、科学技術、制度、情報・メディア、機械システムサービスのような人が作ったものについても該当します。

Trusteeに関わるTrustworthinessの定義にも議論がありますが、Trusteeについては「トラストするに値するかを判断するのに考慮される品質特性」を

指すものとします。信頼性と訳されることが多いのですが、IT分野ではReliability、Dependabilityも信頼性と訳されるので、区別するために「信頼可能性」「信頼相当性」とも訳されます（1 (2) 情報科学を参照）。

Trusteeが多様化することにより、Trustworthinessは人に関わる属性のみならず、機械・システムサービスなどに関する属性も含むこととなります。

(2) Trusteeの複合化の例

情報システムにおいてTrustorであるユーザー視点で見ると、主たるTrusteeは「システムまたはサービス」ですが、単にそれをトラストするだけでなく、それに携わるベンダー、運用者、開発者、保守者などをトラストするかについても複合的に絡んできます。つまり、トラストするかしないかという1つのケースにおいて、相手（Trustee）は1つではなく、複数の相手が複合的に絡むことが多々あります。

(3) 技術の複雑化・自律化

Trusteeが機械・システムサービスなどのケースでは、そこで使われている技術の複雑化や自律化がトラストに影響を与えます。技術の複雑化や自律化は、Trusteeのブラックボックス化、その動作の予測困難化を招くため、Trustorはトラストしにくくなることがあります。

かつては長期間の実績・経験を重ねることでトラストが獲得できたかもしれませんが、今日では技術発展も速く、実績や経験を重ねる十分な期間の確保は難しいと言わざるを得ません。

(4) Trustorの主観

トラストするか否かは、Trustorの主観に最終的には左右されます（3 を参照）。トラストがうまく機能していれば、普段はそれを意識することなく、安心して迅速に行動・意思決定できます。しかし、(1) から (3) が進んでくると、Trustorの主観を左右するさまざまな不確かさが増します。

その結果、トラストが揺らぐと、何もかも心配でとても不安になったり、相手に対する不信感で眠れなくなったりします。深く考えるのはやめてしまったり（常に思考停止）、頼りきり、任せきりになったり、リスクは減少していないのにTrusteeと親密な会話を持つことで信じやすくなり、悪意を持った他者からだまされやすくなることもあります。

頼りきり、任せきりで常に思考停止の状態は好ましいものではなく、トラストすることが無条件に良いことだというわけではありません。

5 おわりに

最後に、トラストの概念と監査の関係に触れておきます。論点は、「監査を含めた保証（アシュアランス）サービスがなぜ存在しているのか？」（存在意義）に関わります。

まず、約束をする当事者ともう一方の当事者との2人が存在していることが、論点の前提です。

すなわち、何らかの約束をする当事者（「約束をする人」）から聞いた内容に賛同し、その人のことを本当に信じたいと思っているもう一方の当事者（「信頼する人」）がいれば、その「約束する人」を信じることはあります。ただし、もう一方の当事者が「約束する人」を信頼できるかどうかは明らかではありません。

「約束する人」（先のモデルのTrusteeに相当）がいて、そこへ「信頼する人（または信頼する可能性のある人。Trustorに相当）」が登場し、「約束する人」が行った約束が「信頼する人」などによって信頼されれば、実際に取引が行われます。すなわち、2人は実際にビジネスその他の社会的な交わりにコミットすることとなります。

次に、もし全ての「約束する人」を信頼でき、約束が信頼できるものであるとすれば、実際、保証を得る必要はありません。言い換えると第三者は必ずしも必要ありません。

以上から、次のことが言えます。すなわち、監査人の地位にある人は実質的に、金融市場における社会的な信頼の創出に中心的な役割を演じています。企業経営陣による誤った、または私利私欲的な、不正が行われているかもしれない財務諸表に脆弱な立場

の人々にコンフォートを供与します。

保証に従事する専門家は、このコンフォートを生産するビジネスに携わっています。同時にコンフォートをも生み出し、公益の守護者でもあります。公益の守護者となり、当然のことながらコンフォートを生むことにより、社会に価値を加える役割を果たすわけですが、状況により保証されることでディスコンフォート（不安）を生み出すこともあり得ます。

ここで実際に利用されるのが保証です。保証は、情報の関連性、信頼性、さらには意思決定者にとってのコンテクスト（文脈、文化の共有度合）を改善することにより、情報の質を向上させる独立したサービスであることが分かっています。この保証により情報の質は向上します。情報の質が向上すれば、「信頼する人」にコンフォートが与えられます。

また、第三者である保証の提供者がいることにより最終的にコンフォートが得られることが理解されれば、保証に対する需要が生じます。すなわち、保証に対する需要が保証を生み出し、それが情報の質を向上させる好循環が生まれます。「信頼する人」にはコンフォートが与えられます。この保証に対する需要が生じ、実際に委託者が監査自体から得られるコンフォートが監査へのさらなる需要を生み出すと考えられます。

今回は「トラスト研究の現場から」として、科学技術振興機構の研究開発戦略センターの最近の学際研究のエッセンスと筆者が考えている箇所について紹介しました。

ここまでお読みいただき気づいた方もいるかと思いますが、現代のトラスト研究は情報科学やAI研究の領域において発展しています。しかし、本研究でも指摘されているとおり、17世紀にまで遡るこれまでのトラスト研究はそれぞれの学問領域、いわゆるディシプリンごとにクローズドな形で行われてきたものであり、相互の知見の共有はほとんど行われていません。

もっとも情報科学やAIにおける成果としての技術は、学問領域を超えてビジネスや会計、監査などにも大きな影響を与えてきていることから、トラスト研究における知見も同様に広く行きわたることが期待されます。

【参考文献】

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（2022）「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/report/CRDS-FY2022-SP-03.html（2023年9月28日アクセス確認）

同（2022）「科学技術未来戦略ワークショップ報告書　トラスト研究戦略　～デジタル社会における新たなトラスト形成～」
https://www.jst.go.jp/crds/report/CRDS-FY2022-WR-05.html（2023年9月28日アクセス確認）

同（2022）「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～」
https://www.jst.go.jp/crds/report/CRDS-FY2021-WR-05.html（2023年9月28日アクセス確認）

同（2023）「公開シンポジウム（2023年1月10日開催）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～」
https://www.jst.go.jp/crds/report/CRDS-FY2022-SY-02.html（2023年9月28日アクセス確認）

Peecher, Mark（2018）Auditing I: Conceptual Foundations of Auditing, Coursera, University of Illinois at Urbana-Champaign

山口 峰男（やまぐち みねお）

PwCあらた有限責任監査法人
PwCあらた基礎研究所　所長
2004年公認会計士登録。「次世代の会計および監査」をテーマとした広範な研究活動に従事。大手銀行において法人融資および本部主計業務に携わったのち、監査法人入所。主に金融機関向けの監査およびアドバイザリー業務に従事し、その後、品質管理本部（金融商品会計、開示、ナレッジマネジメント担当）、英国留学（日本公認会計士協会による大学院派遣）、グローバル教育研修部門（PwC英国にてIFRS金融商品会計の教材開発に従事）などを経て現在に至る。日本証券アナリスト協会認定アナリスト（CMA）（1999年）　日本簿記学会学会賞審査委員（2021年～）
メールアドレス：mineo.yamaguchi@pwc.com