

IT サービスにおける Trustworthiness Trustworthiness on IT Service

神橋 基博[†]
Motohiro Kambashi

1. はじめに

近年における情報技術(IT)の進歩により、人工知能(AI)や IoT といった新しい技術を取り込み、IT サービスを拡大し続けている。このような、IT サービスの拡大は、人材不足の解消、生産性の向上、付加価値の提供といったビジネスにプラスの影響を及ぼす一方で、新たなリスクが生じている。

例えば、AI を用いることで個人に応じたサービスが提供可能になる一方で、個人の属性を分析することでプライバシーの問題、差別や不公平をもたらす恐れがある。

このような問題は新しい技術を用いた IT サービスを導入する上の阻害要因となるものの、情報システムの品質や性能の問題として解決することは困難であり、IT サービスを提供する組織体制、利用者、データといった観点を含める必要がある。

このため、近年では、IT サービスの信頼性(Reliability)に代わり、信頼性(Trustworthiness)に注目が集まっている。

ISO における国際標準化においては、2020 年に AI の Trustworthiness に関する技術文書 ISO/IEC TR 24028:2020^[a] が公開され、2021 年には IoT のシステムライフサイクルに関する国際標準 ISO/IEC 30147:2021^[b] が公開された。

しかしながら、Trustworthiness が意味する内容については、必ずしも明確になっておらず、一般的な信頼性との違いについて混乱が残っている。

本稿では、Trustworthiness に関する従来の議論、および AI の Trustworthiness に関する ISO/IEC TR 24028:2020 および関連する用語の定義を手掛かりに、IT サービスにおける Trustworthiness が意味する内容について考察する。

なお、本稿で参照する ISO における用語の定義は、ISO が提供する Online Browsing Platform (OBP)^[c] および IEC が提供する Electropedia^[d] を参照した。

2. Trustworthiness に関する従来の議論

本章では、一般的な Trustworthiness について、従来の議論を整理する。

2.1 Trust と Trustworthiness

ルーマン(1990)によると、人間の情報処理能力には限界があり、現実世界から得られる情報の全てを処理することができない。そのような不確実性において、相手に対する期待を基に行動するリスクを引き受けることを信頼(Vertrauen/Trust)として定義した。[1]

山岸(1998)は、ルーマンの信頼概念を継承しつつ、信頼を能力に対する期待、意図に対する期待に分類し、意図に対する期待について、長期的なコミットメント関係や相手

の自己利益の評価に基づくものを安心(Assurance)とし、相手の人格や感情に基づくものを信頼(Trust)と定義した。[2]

また、信頼性 (Trustworthiness) を信頼される側の特性、信頼(Trust)を信頼する側の特性と位置づけ、ゲーム理論を用いた心理学的実験から「信頼性」と「信頼」に一貫した関係が無いことを示した。

2.2 Trustworthiness とリスク

中谷内他(2014)は山岸における信頼(Trust)および信頼性(Trustworthiness)の議論を継承しつつ、図 1 に示すように、信頼に影響を及ぼすものとして、信頼する側による信頼される側の「能力」および「動機づけ」に対する認知に加え、信頼する側と信頼される側の価値観の共通性によって説明する主要価値類似性 (Salient Value Similarity : SVS) モデルを導入した。組織に対する信頼および信頼性がどのように変化するかを経年的に調査することで、組織に対する信頼によって認知されるリスクが削減される効果があることを示した。[3]

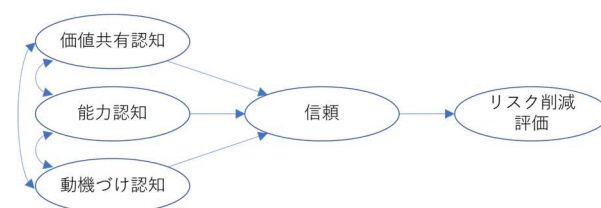


図 1 中谷内他(2014)における信頼モデル[3]

3. ISO における Trustworthiness

3.1 Trustworthiness の定義

ISO/IEC TR 24028:2020 では、Trustworthiness を”ability to meet stakeholders' expectations in a verifiable way”、trust を”degree to which a user or other stakeholder has confidence that a product or system will behave as intended”と定義している。

trust を信頼する側(user or stakeholder)の confidence、trustworthiness を信頼される側の ability として定義している点は、前節の信頼(Trust)および信頼性(Trustworthiness)の関係と整合している。

なお、ISO/IEC TR 24028:2020 では trust と trustworthiness の関係を明示していない。

[a] Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence

[b] Internet of Things (IoT) - Integration of IoT trustworthiness activities in ISO/IEC/IEEE 15288 system engineering processes

[c] <https://www.iso.org/obp/>

[d] <https://www.electropedia.org/>

[†] 情報セキュリティ大学院大学
Institution of Information Security

3.2 Trustworthiness と expectation

Trustworthiness の定義において、expectation は何を意味しているだろうか。

ISO/IEC TR 24028:2020 では定義を示していないものの、ISO では、ISO/IEC Guide 98-4:2012 のように、数学的な期待値とする定義^[e]、および ISO 19208:2016 のように信頼する側の主観として捉える定義^[f]がある。

一方、ISO/IEC TR 24028:2020 における Trustworthiness の定義の補足(Note2)では、Trustworthiness の特徴を示す例として reliability, availability, resilience, security, privacy, safety, accountability, transparency, integrity, authenticity, quality, usability を挙げている。これらの概念は、availability や quality のように計測可能な能力とするもの、accountability や authenticity のように主観によるものを含んでいる。

3.3 Trustworthiness と Verifiable

Trustworthiness の定義において、verifiable とは何を意味しているだろうか。

ISO/IEC TR 24028:2020 では“verification”を“confirmation, through the provision of objective evidence, that specified requirements have been fulfilled”と定義している。客観的な証拠を前提としている点は、前節における availability^[g]や quality^[h]といった数学的な期待値の親和性が高い一方、accountability^[i]や authenticity^[j]といった主観による概念に対しては、客観的な証拠をどうやって確保するのかという点が課題となる。

3.4 Trustworthiness と Reliability

3.2 で示した通り、Trustworthiness の特徴の一つとして Reliability が挙げられている。一方、JIS Z 8115-2019 では「信頼性」に“Reliability”を対応させる用例が認められる。当該規格は IEC 60050-192 及び Amendment 1:2016 を基にしており、IEC では能力として定義^[k]されている。

以上より、Trustworthiness は Reliability における能力概念を含む、より広義な概念であると言える。

[e] “for a continuous random variable X characterized by a PDF $g_x(\xi)$ ”

[f] “societal expectations”について、“what society deems as acceptable”と定義。(以下、下線部は筆者による)

[g] ISO/IEC 27000:2018 では“property of being accessible and usable on demand by an authorized entity”と定義。

[h] ISO 20205:2015 では“quality of service”として“The ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow.”と定義。

[i] ISO/IEC TR 24028:2020 では“property that ensures that the actions of an entity may be traced uniquely to that entity”と定義。

[j] ISO/IEC 27000:2018 では“property that an entity is what it claims to be”と定義。

[k] “ability to perform as required, without failure, for a given time interval, under given conditions”

4. 考察

従来の議論では、信頼性(Trustworthiness)を能力に対する期待(認知)、意図に対する期待(認知)の2つ、あるいは価値共有を加えた3つの分類していたのに対して、ISO/IEC TR 24028:2020 では、明確な区別を示していない。

定義の補足に能力以外の概念が含まれていること、また、能力に関する期待という側面においては、Reliability と同義となることを考慮すると、ISO が Trustworthiness をとりあげる目的には、Reliability では捉えられない意図あるいは価値共有に関する期待を必要としていたと考えられる。

ハードウェア、ソフトウェア、ネットワーク、データ等から構成される IT サービスには意図も価値共有も無いことから、Trustworthiness が議論される背景には、IT サービスを提供する組織、あるいは利用者などを含んだエコシステムといったものが想定される。

その一方で、定義に“verifiable”が含まれており、いることから、意図および価値共有に関する Trustworthiness について、どのように客観的な証拠を示すのかが課題となる。

5. まとめ

本稿では、Trustworthiness に関する従来の議論と ISO における定義から、Trustworthiness が意味する内容について考察した。

社会が機能するために信頼(Trust)が不可欠であることは言うまでもないが、信頼を向上させるためには、信頼される側の信頼性(Trustworthiness)が必要となる。ISO/IEC TR 24028:2020 における定義は、能力に対する期待を前提とする Reliability を包含しているだけでなく、意図に対する期待という観点を含んでいる。

Reliability から Trustworthiness へと信頼性概念を拡大する一方、信頼性をどのように検証するか(Verifiable)という点について、ISO の定義は十分ではない。今後の検討の動向を注視する必要がある。

なお、信頼および信頼性については、様々な議論があり、本稿で紹介したルーマン、山岸、中谷内らの議論が唯一の解ではない。また、ISO および IEC においても、異なる標準において Trustworthiness および関連する用語の定義が共通であるという保証は無い点には留意が必要である。

その上で、ISO/IEC TR 24028:2020 および関連する標準を用いて IT サービスあるいは AI、IoT 等に関する信頼性を議論する際には、意図や価値観といったものに対する信頼性をどのような証拠をもって評価するのかという点に関する議論が必要であり、本稿で検討した内容を踏まえ、引き続き動向の把握に努めたい。

参考文献

- [1] 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." 勁草書房 (1990).
- [2] 山岸俊男. 信頼の構造. 東京大学出版会, (1998)
- [3] 中谷内一也, 工藤大介, and 尾崎拓. "東日本大震災のリスクに深く関連した組織への信頼." 心理学研究 85.2 (2014): 139-147.