検証可能な信頼へ
～デジタル社会におけるトラスト形成を考える～

セコム株式会社　IS研究所
島岡政基

本日お伝えしたいこと

- ✓ デジタル社会では、アナログ文化に依拠した旧来のトラストから、検証可能なトラストへの転換が必要
- ✓ 高度化した技術の存在を前提とした、デジタル社会のあるべきトラストを考える必要
- ✓ フィジカルの証拠をサイバーへ写像し、“継続的に”状態を確かめ続ける世界へ

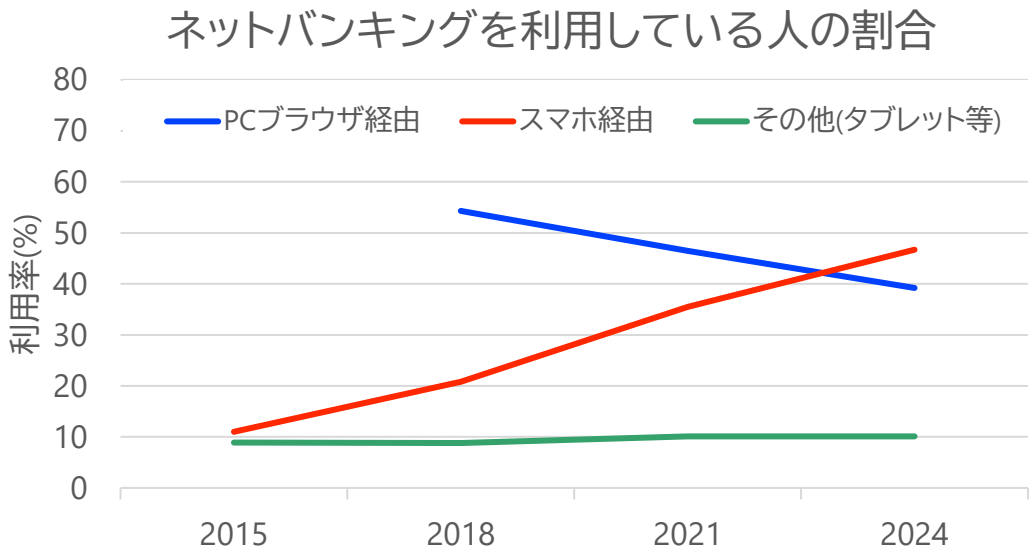
- **Part1：トラストの本質と「検証可能性」**
- **Part2：ハードウェアセキュリティと検証可能性**
- **Part3：トラスト研究の多様なアプローチ**

Part 1：
トラストの本質と「検証可能性」

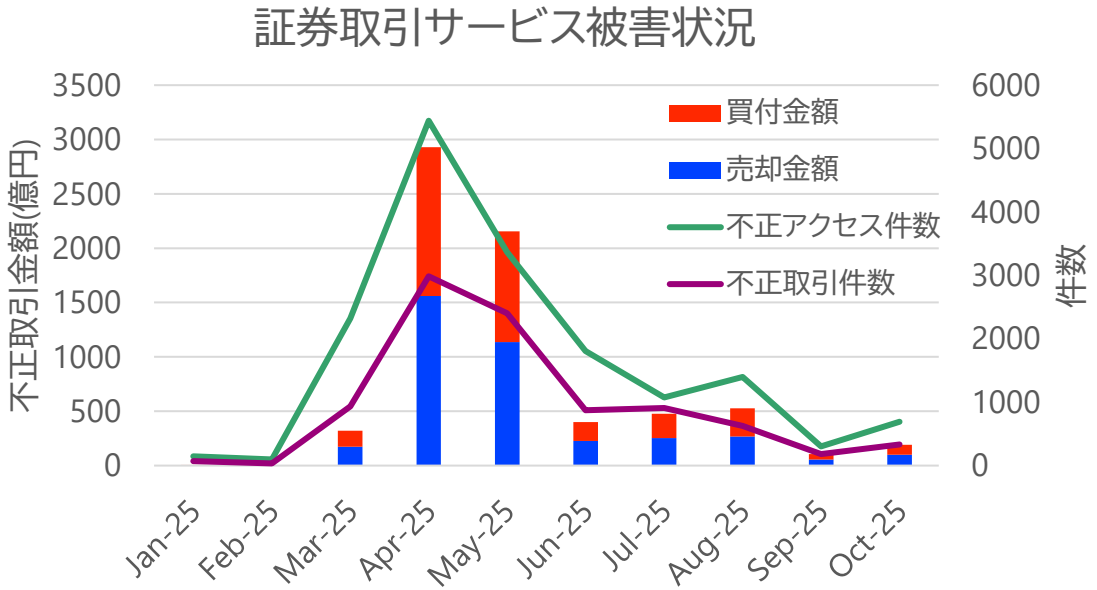
アナログ社会のトラストから、
検証可能性にもとづくデジタル社会のトラストへ

デジタル化で変わる日常とトラストの課題

- デジタルサービスの普及により、端末・ネットワーク経由でのサービス利用が日常化
  - 店舗での対面サービスから「ログイン」「API経由取引」「モバイル／Web決済」などへ
- 「誰と取引しているのか」「何が行われているのか」の不透明化
  - 従来の物理的・対面型リスク（書類の偽造や改ざん）
    → デジタル特有のリスクへと変化（なりすまし、生成AI詐欺、中間者攻撃、改ざんなど）



出典：全国銀行協会「よりよい銀行づくりのための アンケート報告書」(2024)，第2章(p.14)の数値をもとに作図



出典：金融庁「インターネット取引サービスへの不正アクセス・不正取引による被害が急増しています」(2025年11月更新) の数値をもとに作図

デジタルの急速な発展・普及で利便性が高まる一方、トラストの課題も浮き彫りになってきた

アナログ社会：顔の見える関係の寄与

私たちが無意識に依存していた「トラストの仕組み」は、
デジタル化によって前提が崩れている。

- 対面効果や対物証拠（書類・印鑑・署名）を前提とした法制度や慣習
  - 物理的書類の改ざん困難性
  - 対面での非言語コミュニケーション(仕草、視線、表情、”間”など)とだましにくさ
- アナログ社会のトラストの多くは、無意識にこれらに依存する形で形成されてきた

図版生成の一部に生成AIを利用しています

デジタル社会:顔の見えない相手と、見えない仕組み

- デジタル技術の発展と普及は社会に恩恵をもたらす一方で、お互いの顔が見えない関係性や、システムの高度・複雑化などによる弊害も出始めている

例1) なりすましの多様化:

- ID／パスワード盗用だけでなく、生成AIによる音声／映像偽装、ディープフェイク、ソーシャルエンジニアリングの高度化

免許証の顔写真をディープフェイクで自身の写真とすり替えてオンライン審査を潜り抜け、クレジットカードを契約し、キャッシング機能で現金を詐取。
生成AI偽画像で「本人なりすまし」、口座開設デジタル顔認証すり抜け…闇サイトでの手口公開にコメント続々：読売新聞

例2) システム内部のブラックボックス化:

- AI／機械学習モデルの「ブラックボックス性」
- クラウドサービス、マイクロサービス、コンテナ、API経由の処理／通信チャネルの複雑化

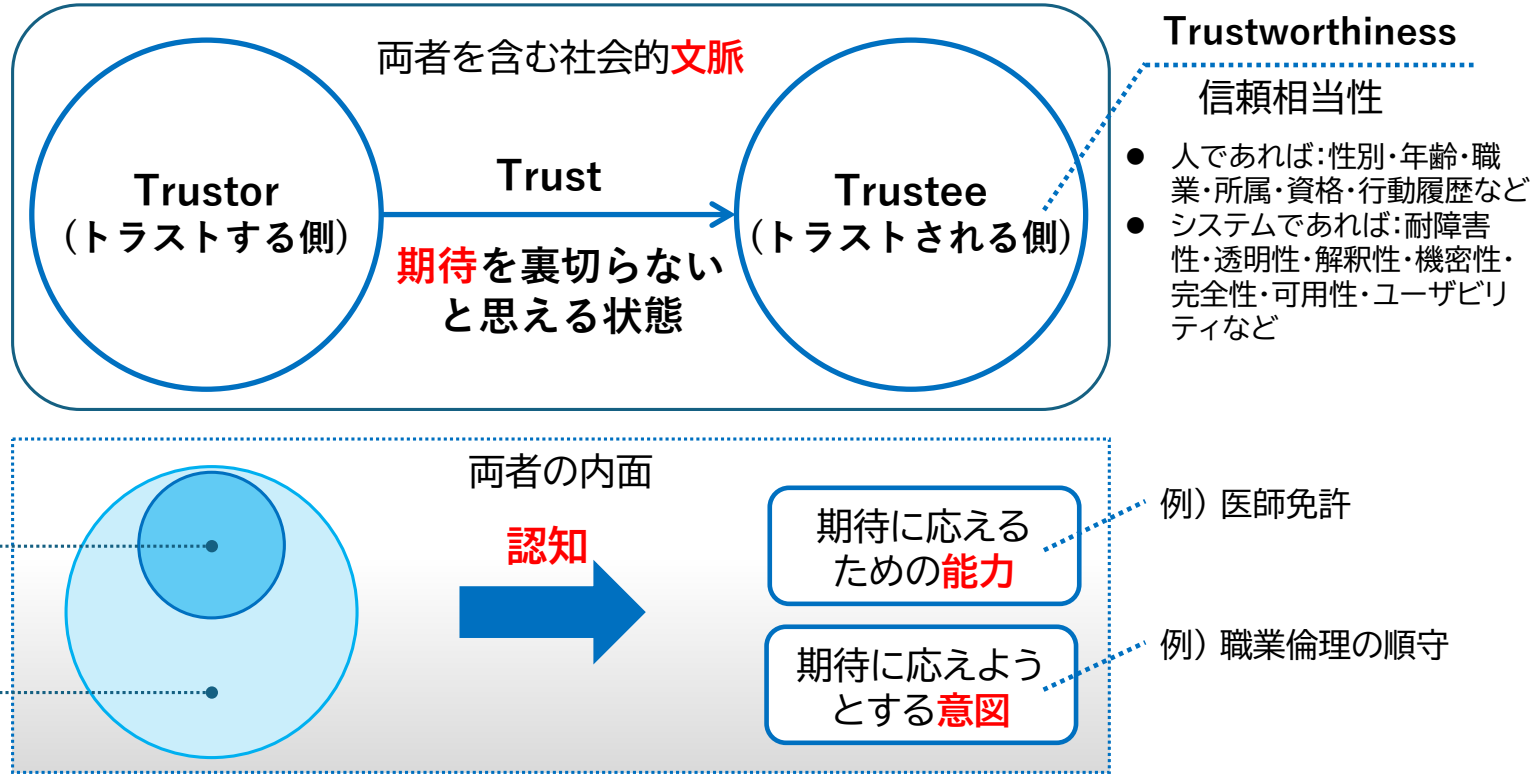
2010年5月、米株式市場でダウ平均がわずか数分で9%（約1000ドル）下落。先物価格の急落を機に高頻度取引(HFT)などのアルゴリズム取引が複雑に連動してしまい、人間による介入が間に合わなかった。
複雑化した米国の株式市場構造と「フラッシュ・クラッシュ」：野村総合研究所

デジタル化が進む中で、人間系による判断・制御は限界を迎えている

特に異常や不正に関連する要素を人間系だけに依拠し続けることは、
系全体の脆弱性を高めることにもつながりかねない

高度で複雑な技術の存在を前提としたトラストの再構築が必要

トラストを構成する5つの要素



期待： トラストする側（Trustor）が、相手（Trustee）に「自分にとって良い行動をしてくれる」と期待すること

能力： Trusteeがその期待に応えるための力や特性

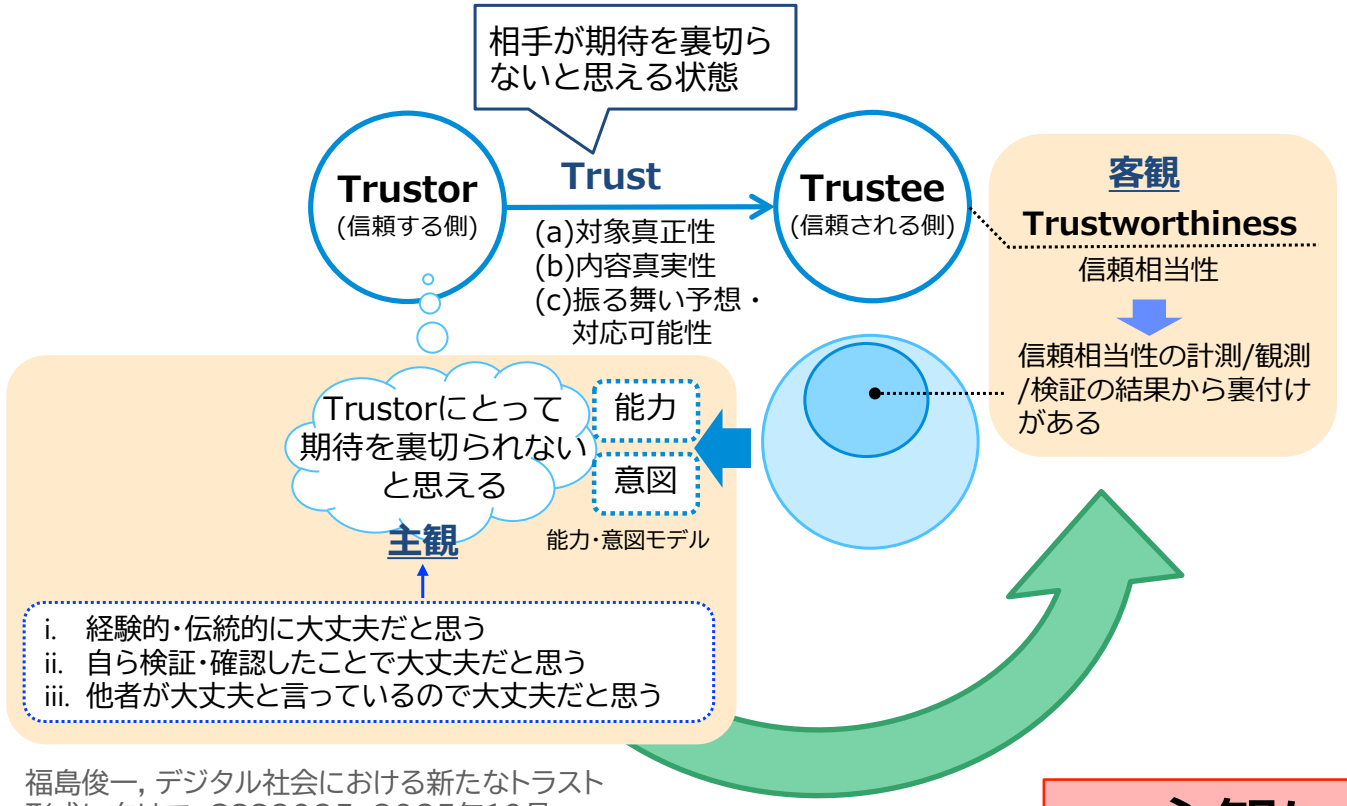
意図： Trusteeが期待に応えようとする誠実さや善意

認知： TrustorがTrusteeの能力や意図をどう認識しているか

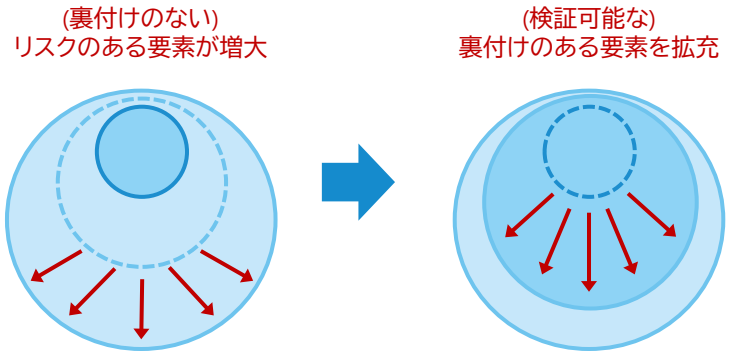
文脈： 両者の関係性や状況

トラストにおける、主観と客観の二面性

- トラストは、客観的な情報を参照しつつも主観的な判断を伴う、ある種不合理な仕組み



福島俊一, デジタル社会における新たなトラスト形成に向けて, CSS2025, 2025年10月.



主観に依拠してきた要素を如何に客観的要素に変えていけるかがポイント

デジタル社会の不確実性とトラストの3側面

- デジタル空間において、アナログ社会よりも不確実性が高まる3つの側面

戦略プロポーザル：デジタル社会における新たなトラスト形成, CRDS-FY2022-SP-03,
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, 図5-1より, p.35, 2022年9月.

トラストの3側面	現状の社会的よりどころ	脅威	デジタル社会の課題
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	なりすまし、改ざん	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、学説は査読性による学術コミュニティ合意など	ねつ造、誤記等	AIによるフェイク生成が高品質化したため、写真・映像の証拠性が揺らぎつつある。そもそも絶対的真実・事実は定まらず、ファクトチェック可能な対象は限定的。
振舞い予想・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	内部不正、悪意あるコードなど	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。

技術のみならず社会制度・慣習等により複合的に形成されてきた。
普段は無意識・暗黙裡によりどころとしているものも少なくない。

これらの検証可能性を
どう実現していくべきか？

検証可能性：裏付けにもとづいてトラストする仕組み

- 客観的証拠にもとづいて第三者が画一的に真偽を判定できる能力

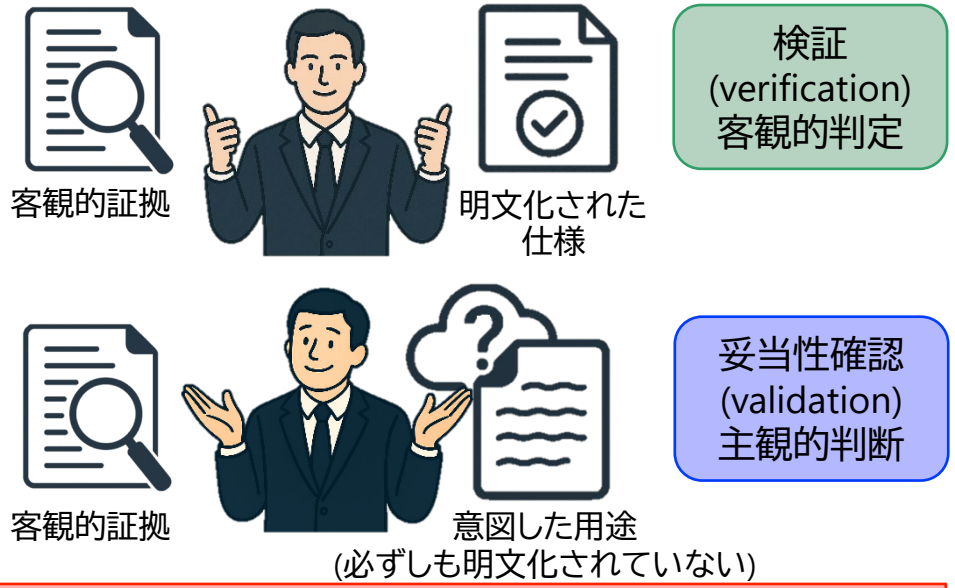
- 暗黙的・主観的な要素を排除し、誰が確認しても同じ結果が期待できるようにする。
- デジタル社会では、証拠の電子化、処理のブラックボックス化が進むため、第三者による検証の重要性が高まる。

【V&V問題】狭義にはVerificationとValidationは明確に定義の違いがあるが、一般的には両者含めて扱われることが多い

検証（verification）：明文化された仕様に適合していることを、客観的証拠を用いて確認すること

妥当性確認（validation）：**意図した用途**に適していることを、客観的証拠を用いて確認すること

— ISO 9000:2015(意訳)



検証可能性を高めることで機械処理を可能とし、
人間系の介入（妥当性確認）を必要とする範囲を最小化する

図版生成の一部に生成AIを利活用しています

小括：なぜ検証可能なトラストが必要なのか

- 不確実性の増大
- 人間系に依拠したトラストの限界
- 検証可能なトラストの必要性
  - 観測するための技術
  - 判定するための基準
  - 自律するための主観の明文化

} Part 2で解説していきます

Part 2：
ハードウェアセキュリティと検証可能性

検証可能性の物理的な土台としてのハードウェアセキュリティ

トラストの3側面と検証可能性の具体化

- トラストの3側面について、デジタル社会でそれぞれの検証可能性を具体化するには何が必要か
- 技術に落とし込むことで機械化・自動化を進め、人間系への依存を最小化していく

トラストの3側面	検証したいこと	必要となる証拠
対象真正性 誰・何であるかの確かさ	対象は本当に相手の主張している通りの存在か？	・本人・機器などの識別情報 ・なりすまし防止のための認証情報(証明書や電子署名等)
内容真実性 内容が正しいか	「どの内容を」「どの手順で」「誰が測定・取得したか」の透明性	・作成元データ等 ・作成手続き、評価手続き等の記録情報
予想・対応可能性 対象の振る舞いをどの程度予想できるか	システムの振る舞いをどの程度の確度で予想できるか(継続的な観測)	・過去の動作記録とその時系列変化 ・状態の変化を追跡し得る情報(イベントログ等)

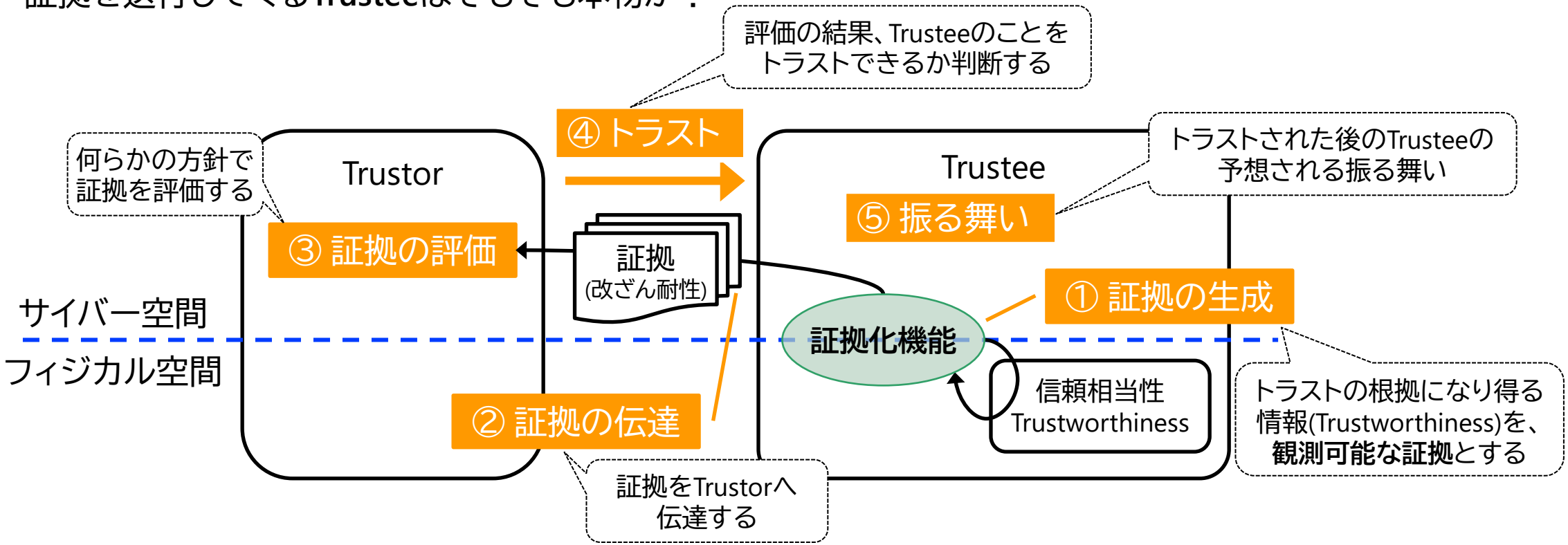
各証拠を正しく収集し、検証可能性を備えた形で
Trustorに伝えるための仕組みとその安全性確保が必要

図版生成の一部に生成AIを利用しています

フィジカル空間の証拠をサイバー空間へ

- デジタル社会では、実空間から収集した証拠を、サイバー空間で検証できるように写像する仕組みが不可欠

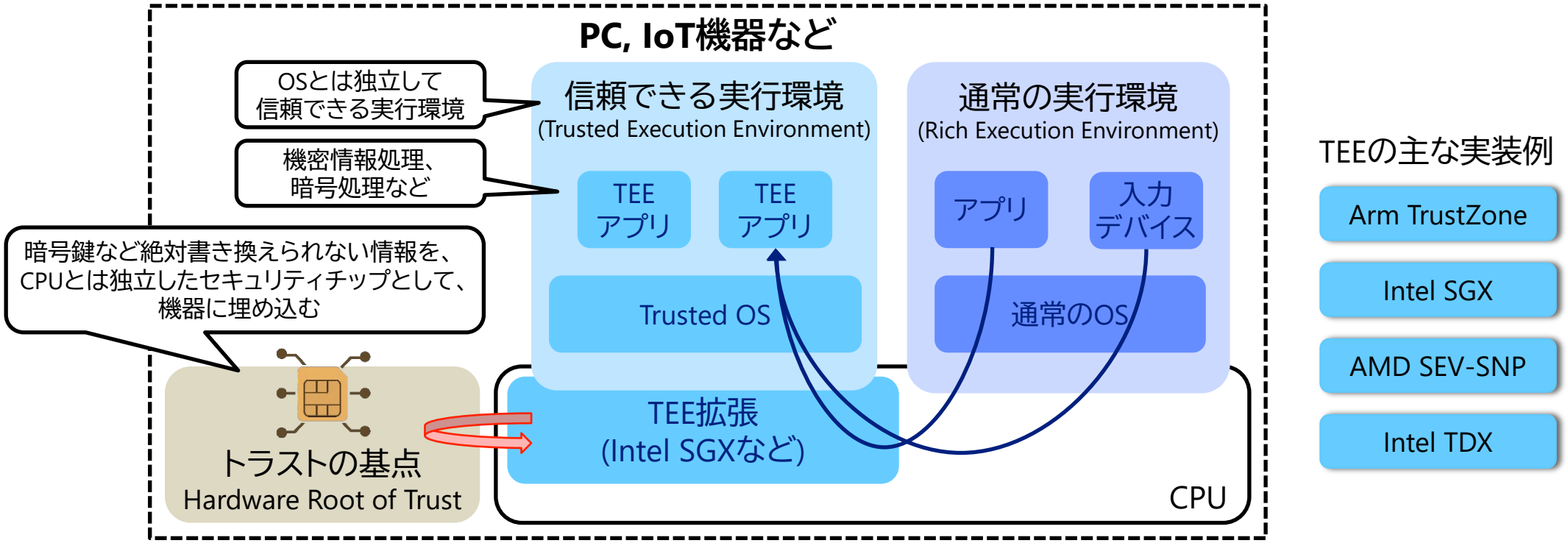
- 実空間の情報を、サイバー空間で観測可能な情報とするには？
- サイバー空間を介して安全にTrustorへ伝達するには？
- 証拠を送付してくるTrusteeはそもそも本物か？



Trusted Execution Environment(TEE)とRoot of Trust(RoT)

- TEE：CPU内に隔離された、信頼された（安全な）実行環境
  - 通常のOSとは別のメモリ空間やAPIを持ち、高い安全性を提供する。
  - モバイル決済、生体認証、暗号鍵管理、秘密計算などの暗号処理や機密情報の保護に使われ、リモートアテステーションの基盤となる。
- RoT：CPUより下位でブートローダーの安全性や機器外部への署名を保証する仕組み

安全性を確保すべき領域を、より最小化・隔離する流れ



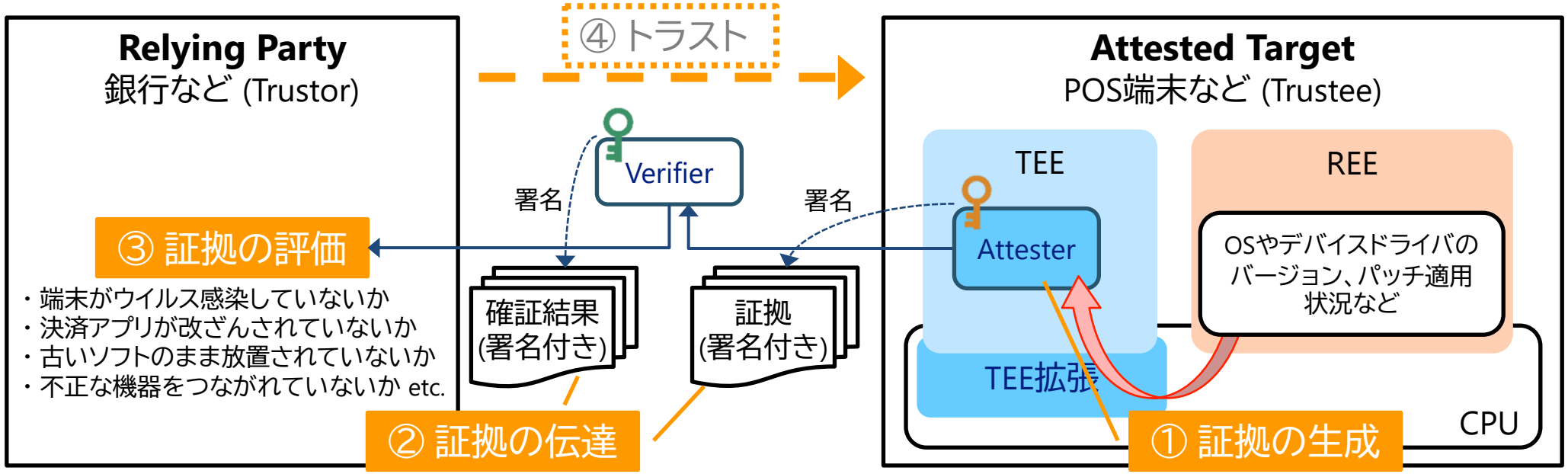
TEEの主な実装例

- Arm TrustZone
- Intel SGX
- AMD SEV-SNP
- Intel TDX

図版生成の一部に生成AIを利用しています

Remote Attestation

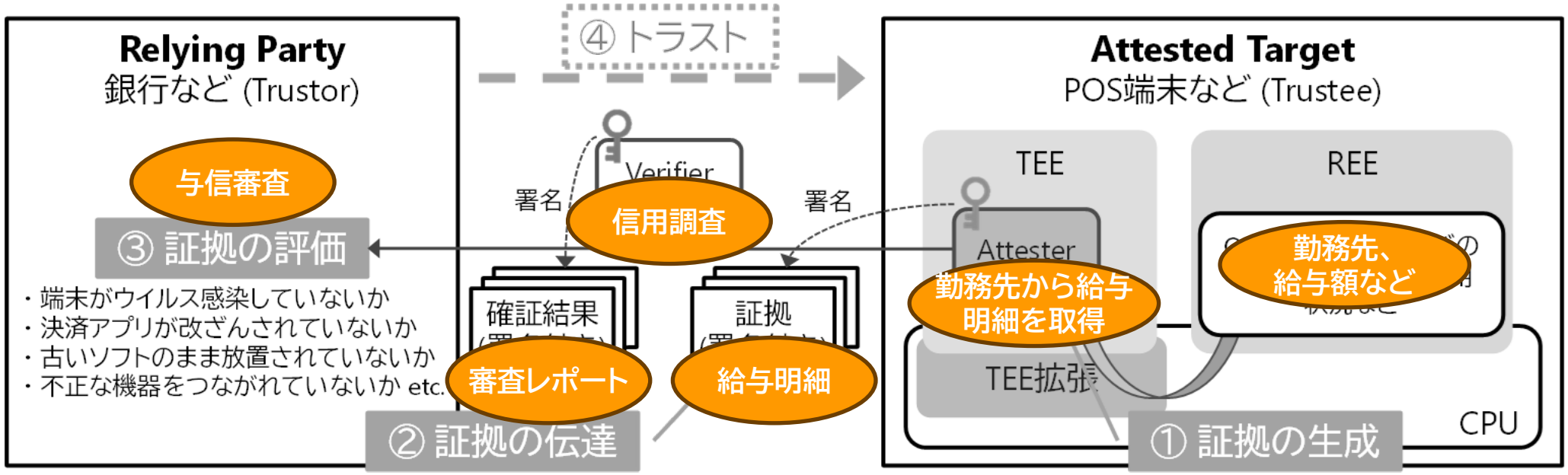
- 外部(遠隔)から、デバイスの状態を証拠として下記を検証する仕組み
  - デバイスが**改ざんされていないこと**
  - 実行中のソフトや構成が**期待どおりであること**
  - 信頼できるハードウェア基盤(TEE)が**証拠の真正性を担保**していること
- 「相手が本当に信頼できる状態で動作している」ことを証明するための基盤として機能する



図版生成の一部に生成AIを利活用しています

Remote Attestation

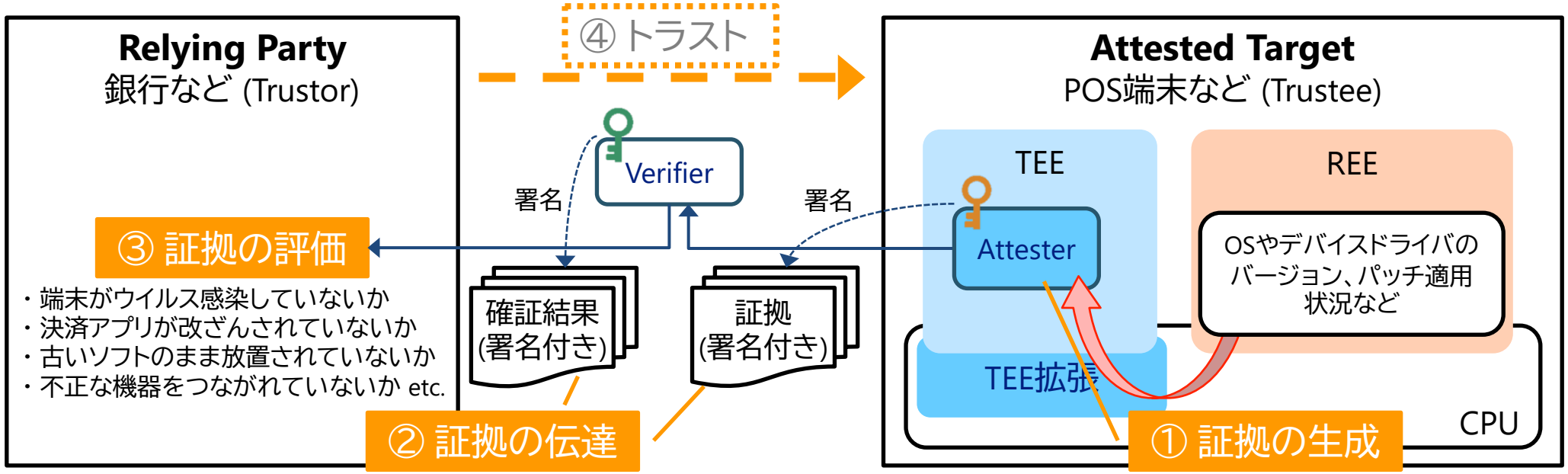
- 外部（遠隔）から、デバイスの状態を証拠として下記を検証する仕組み
  - デバイスが**改ざんされていないこと**
  - 実行中のソフトや構成が**期待どおりであること**
  - 信頼できるハードウェア基盤（TEE）が**証拠の真正性を担保**していること
- 「相手が本当に信頼できる状態で動作している」ことを証明するための基盤として機能する



図版生成の一部に生成AIを利用しています

Remote Attestation

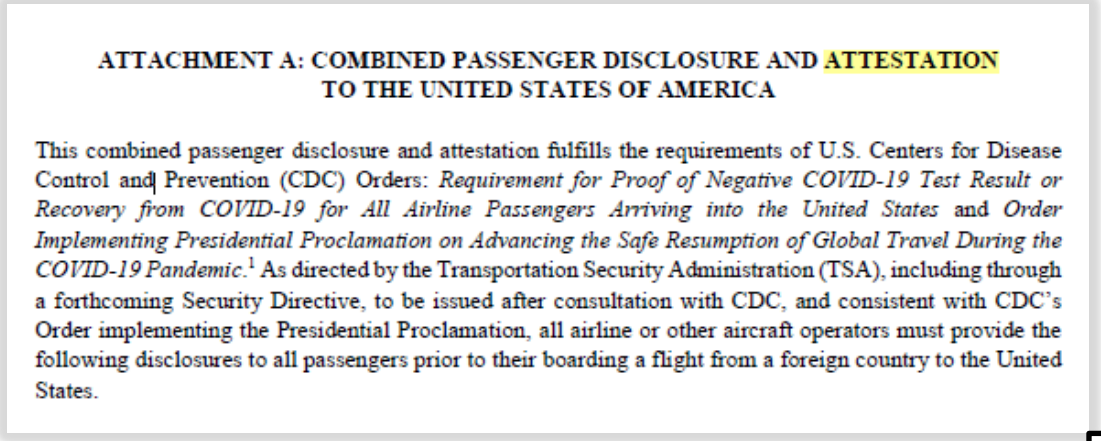
- 外部(遠隔)から、デバイスの状態を証拠として下記を検証する仕組み
  - デバイスが**改ざんされていないこと**
  - 実行中のソフトや構成が**期待どおりであること**
  - 信頼できるハードウェア基盤(TEE)が**証拠の真正性を担保**していること
- 「相手が本当に信頼できる状態で動作している」ことを証明するための基盤として機能する



図版生成の一部に生成AIを利活用しています

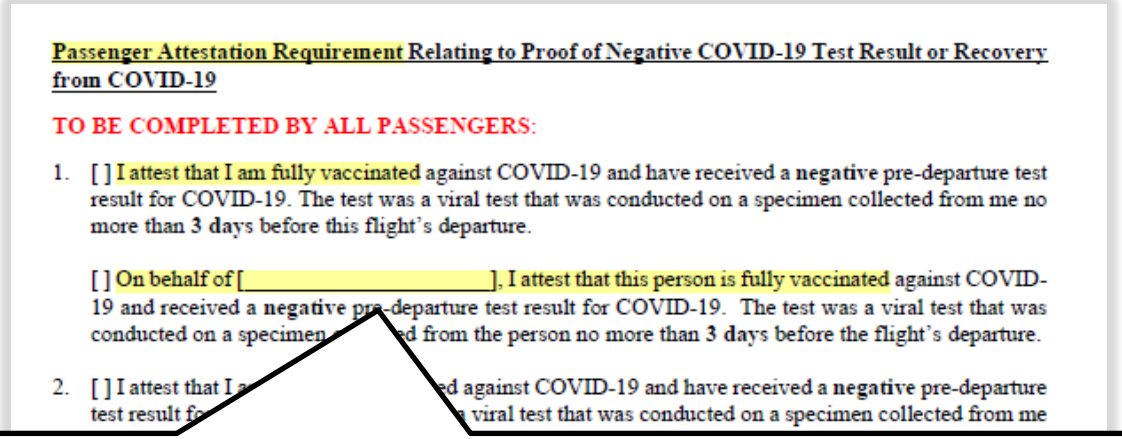
Attestationとは

- 自分が見聞きしたことに責任を以て保証・証言すること（意訳）
  - to show or prove that something is true (ロングマン現代英英辞典)
  - Declare（宣誓）と異なり明確に法的責任を負う行為
  - Verificationは、これらを証拠化し、明文化された仕様に適合していることを確認すること



ATTACHMENT A: COMBINED PASSENGER DISCLOSURE AND ATTESTATION TO THE UNITED STATES OF AMERICA

This combined passenger disclosure and attestation fulfills the requirements of U.S. Centers for Disease Control and Prevention (CDC) Orders: *Requirement for Proof of Negative COVID-19 Test Result or Recovery from COVID-19 for All Airline Passengers Arriving into the United States* and *Order Implementing Presidential Proclamation on Advancing the Safe Resumption of Global Travel During the COVID-19 Pandemic*.[1] As directed by the Transportation Security Administration (TSA), including through a forthcoming Security Directive, to be issued after consultation with CDC, and consistent with CDC's Order implementing the Presidential Proclamation, all airline or other aircraft operators must provide the following disclosures to all passengers prior to their boarding a flight from a foreign country to the United States.



Passenger Attestation Requirement Relating to Proof of Negative COVID-19 Test Result or Recovery from COVID-19

TO BE COMPLETED BY ALL PASSENGERS:

1. [ ] I attest that I am fully vaccinated against COVID-19 and have received a **negative** pre-departure test result for COVID-19. The test was a viral test that was conducted on a specimen collected from me no more than **3 days** before this flight's departure.

   [ ] On behalf of [_____], I attest that this person is fully vaccinated against COVID-19 and received a **negative** pre-departure test result for COVID-19. The test was a viral test that was conducted on a specimen collected from the person no more than **3 days** before the flight's departure.

2. [ ] I attest that I a... ed against COVID-19 and have received a **negative** pre-departure test result fo... viral test that was conducted on a specimen collected from me

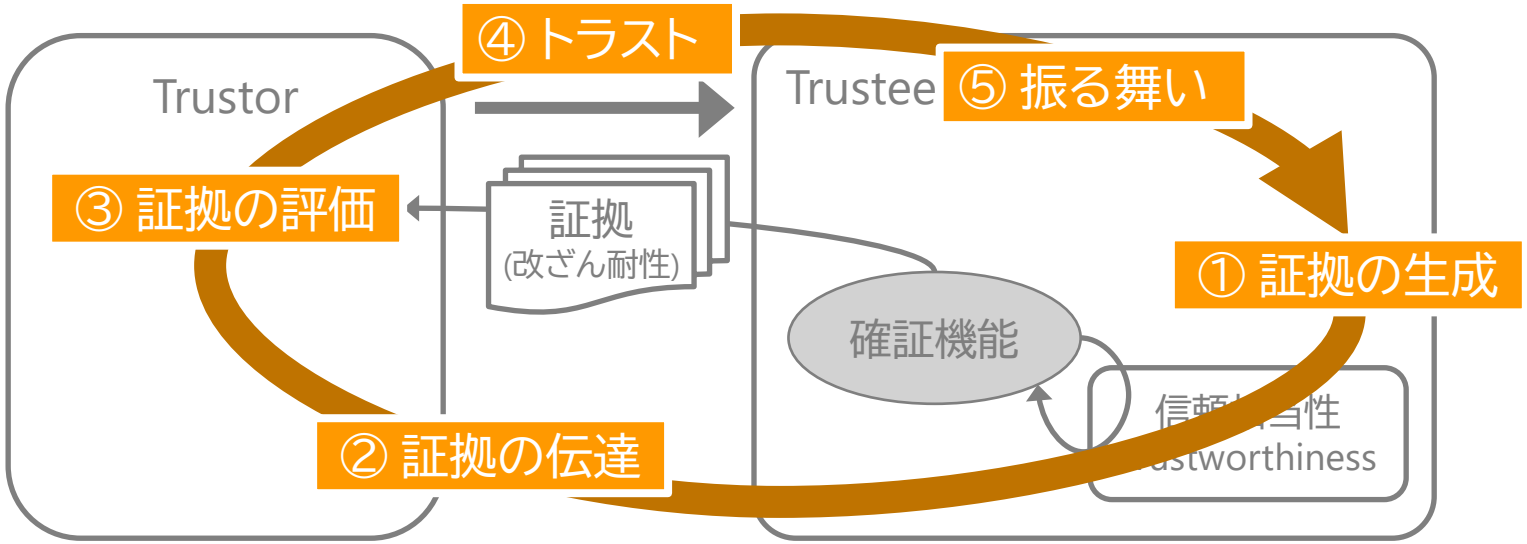
- 米国行き航空便に搭乗する際に乗客が航空会社に提出する申告書類の添付様式A(既に廃止)
- COVID-19検査結果に関する証言要件
- 未成年などの場合は法定代理人がattestする

Passenger Attestation Requirement Relating to Proof of Negative COVID-19 Test Result or Recovery from COVID-19
（中略）
1. [ ] I attest that I am fully vaccinated against COVID-19（中略）
   [ ] On behalf of [_____], I attest that this person is fully vaccinated against COVID-19（中略）

出典：https://www.ana.co.jp/topics/coronavirus-travel-information/pdf/CDC_Attestation_Form.pdf

ToBe：継続的・自律的なアテステーション

- 一時的な安全確認ではなく、“継続的にトラストを確かめ続ける仕組み”へ
  - 変化し続けるシステム環境においては、一度だけのアテステーションでは不十分で、状態変化に追随しながら常に“今この瞬間の健全性”を示すことが不可欠となる。

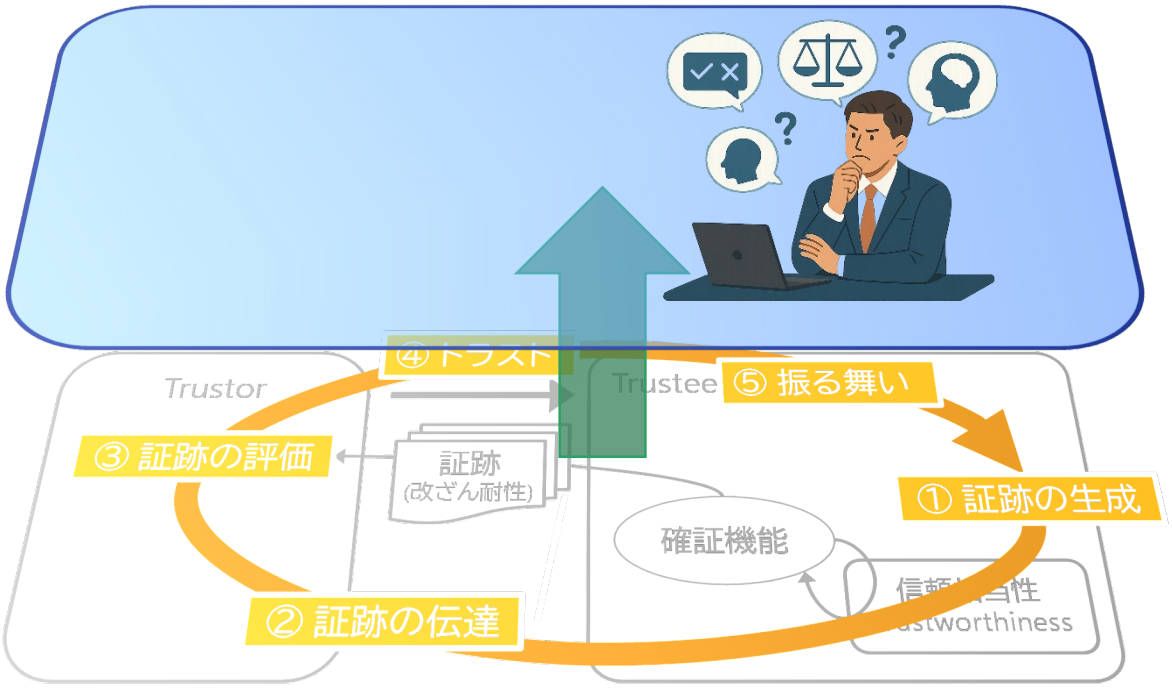


- システムの状態は常に変化し、単発の検証では変化後の不正を捕捉できない。
- 自律的なサービス連携には、状況に応じて自動かつ随時のトラスト評価が求められる。
- 継続検証により潜伏期間を最小化し、「常に見られている」状況を抑止につなげる。

スケーラブルで持続可能なトラストへ向けて

- 高頻度の検証は機械が、低頻度の最終判断（妥当性確認）は人が行う

- 妥当性確認を別レイヤとすることで自律・短周期での検証が可能になる
- 人間系の介入頻度を最小限にとどめることで、丁寧かつ慎重な判断力を確保する



Pros	Cons
原因特定が容易 検証の問題か、判断の問題かが分離され、誤検知や不具合の「どこが悪いか」を迅速に絞り込める。	**判断の分断リスク** 検証結果の意味を現場が理解できない、あるいは業務文脈が検証ロジックに反映されないなど、判断が分断されるリスクがある。
人の判断リソースを節約 客観レイヤが正常系を自動処理し、異常時のみ主観レイヤに上げるため、人の判断リソースが節約される。	**エスカレーションによる遅延** 主観レイヤへのエスカレーションを経るため、リアルタイム対応が必要な領域では遅延が問題になる場合がある。
ガバナンスの明確化 「機械が保証する範囲」と「人が納得すべき範囲」がはっきり説明でき、監査・説明責任に強い構造になる。	**境界設定の難しさ** どの状態で主観レイヤに上げるのか、そのしきい値や条件の設計が難しく、誤判定や過剰エスカレーションの要因になる。

図版生成の一部に生成AIを利用しています

小括:ハードウェアセキュリティで実現する検証可能性

- フィジカル空間とサイバー空間をつなぐハードウェアセキュリティ
- ハードウェアセキュリティを基盤とするTEEとRemote Attestation
- トラストをスケーラブルにする継続的・自律的なアテステーション

Part 3：
トラスト研究におけるアプローチの多様化

技術・制度・社会の総合知に基づく取り組み
客観と主観の両面的アプローチ

トラスト研究開発の目指す姿

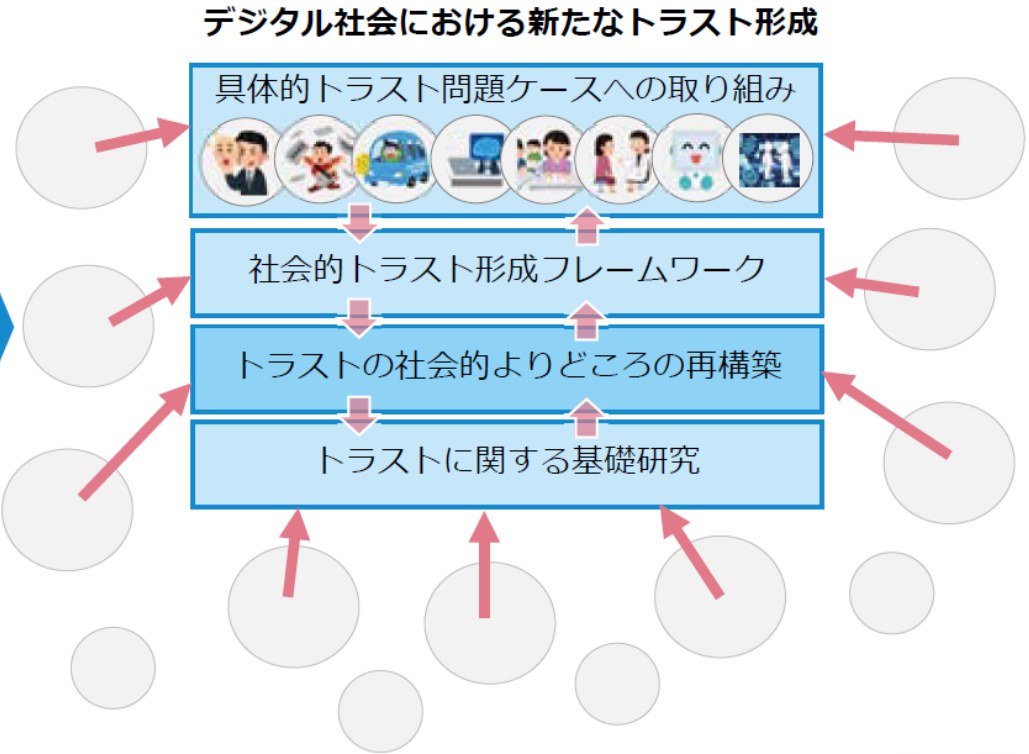
- 分野個別の一面的な研究(とその切り貼り)では不十分であり、技術開発だけでなく制度設計や社会受容も考慮に入れた**総合知**に基づく取り組みが必要

Before

さまざまな分野でトラストに関わる研究が実施されているが、それらの間で知見共有・連携はほとんどなく、それぞれはトラスト問題に対して個別的な対処、断片的な状況改善にとどまる

After

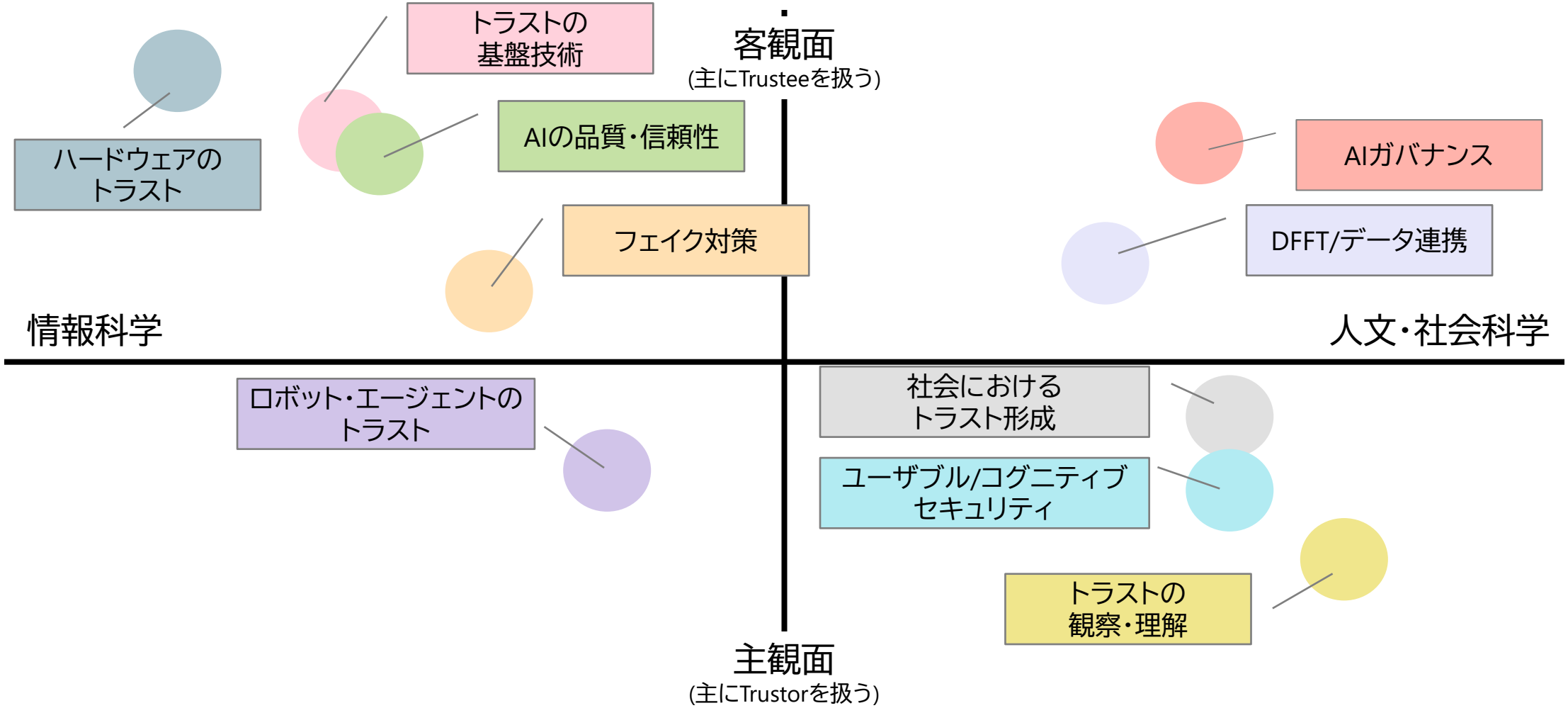
デジタル社会における新しいトラストの仕組みとそれによるトラスト問題対策の全体ビジョンを描いて共有し、具体的トラスト問題と共通基礎の両面から連携して、社会に貢献する研究を目指す



戦略プロポーザル：デジタル社会における新たなトラスト形成, CRDS-FY2022-SP-03, 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, 図1-2, p.3, 2022年9月.

トラスト研究におけるアプローチの多様化

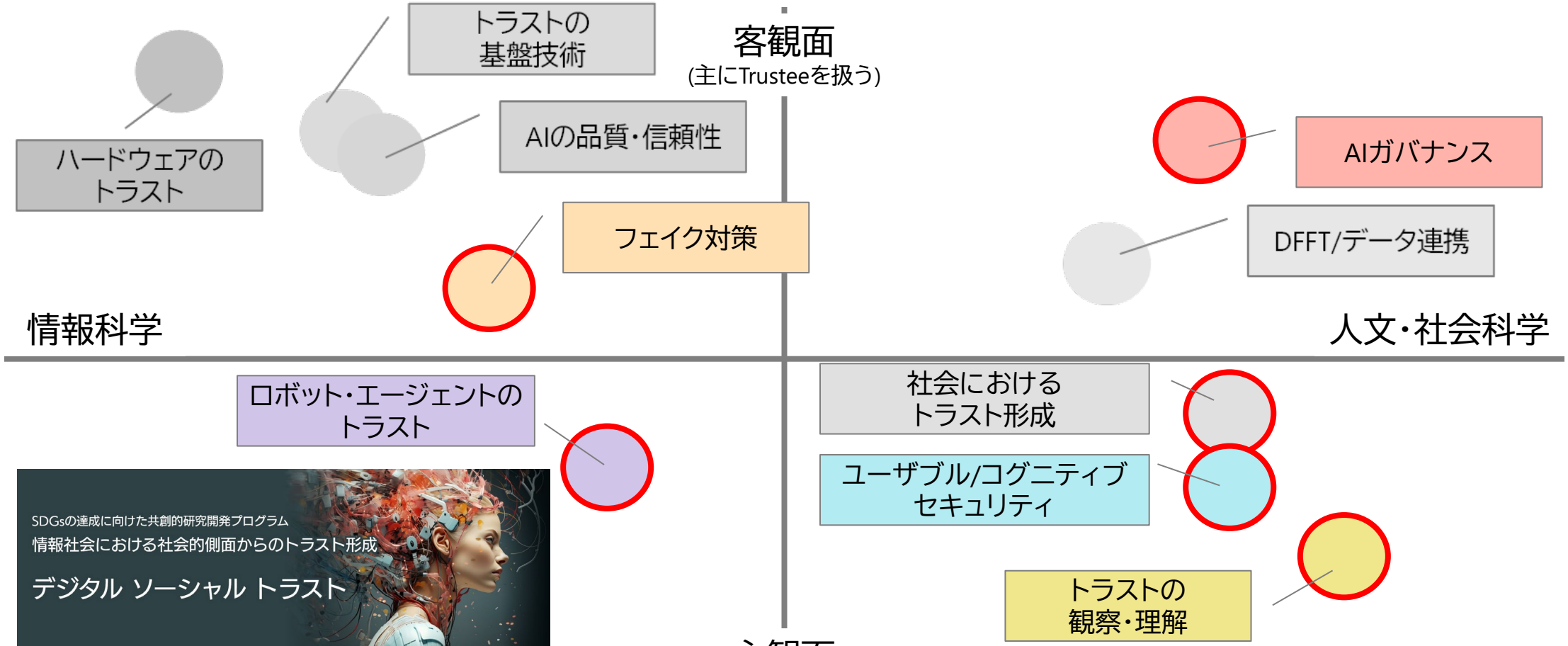
- 近年のトラスト研究および関連分野について、情報科学と人文・社会科学、客観面と主観面から整理（各分野の詳細は後述）。



図版生成の一部に生成AIを利活用しています

RISTEXデジタルソーシャルトラストプログラム

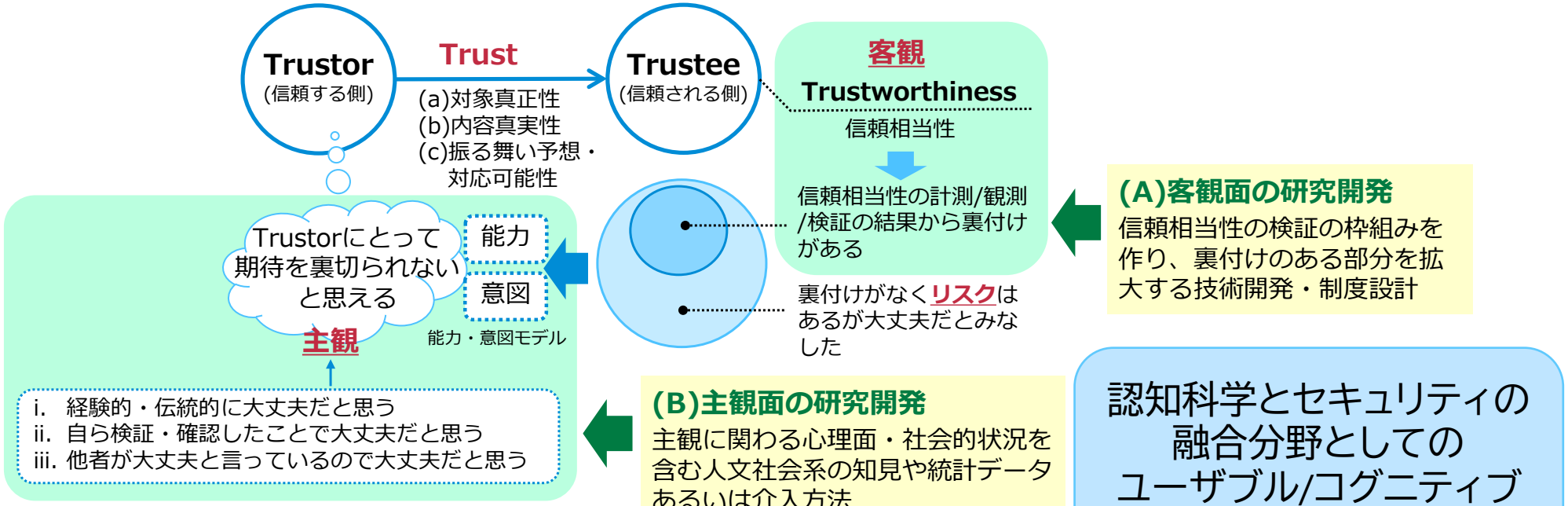
規制、経済、技術活用、教育といった多面的な観点からのアプローチや
多様な主体の連携を図りながら総合知の活用を進めていく



https://www.jst.go.jp/ristex/digist/projects/
各プロジェクトの情報をもとに演者が独自に分類したもの

図版生成の一部に生成AIを利活用しています

トラストの客観面と主観面の両面的アプローチ



福島俊一, デジタル社会における新たなトラスト形成に向けて, CSS2025, 2025年10月.

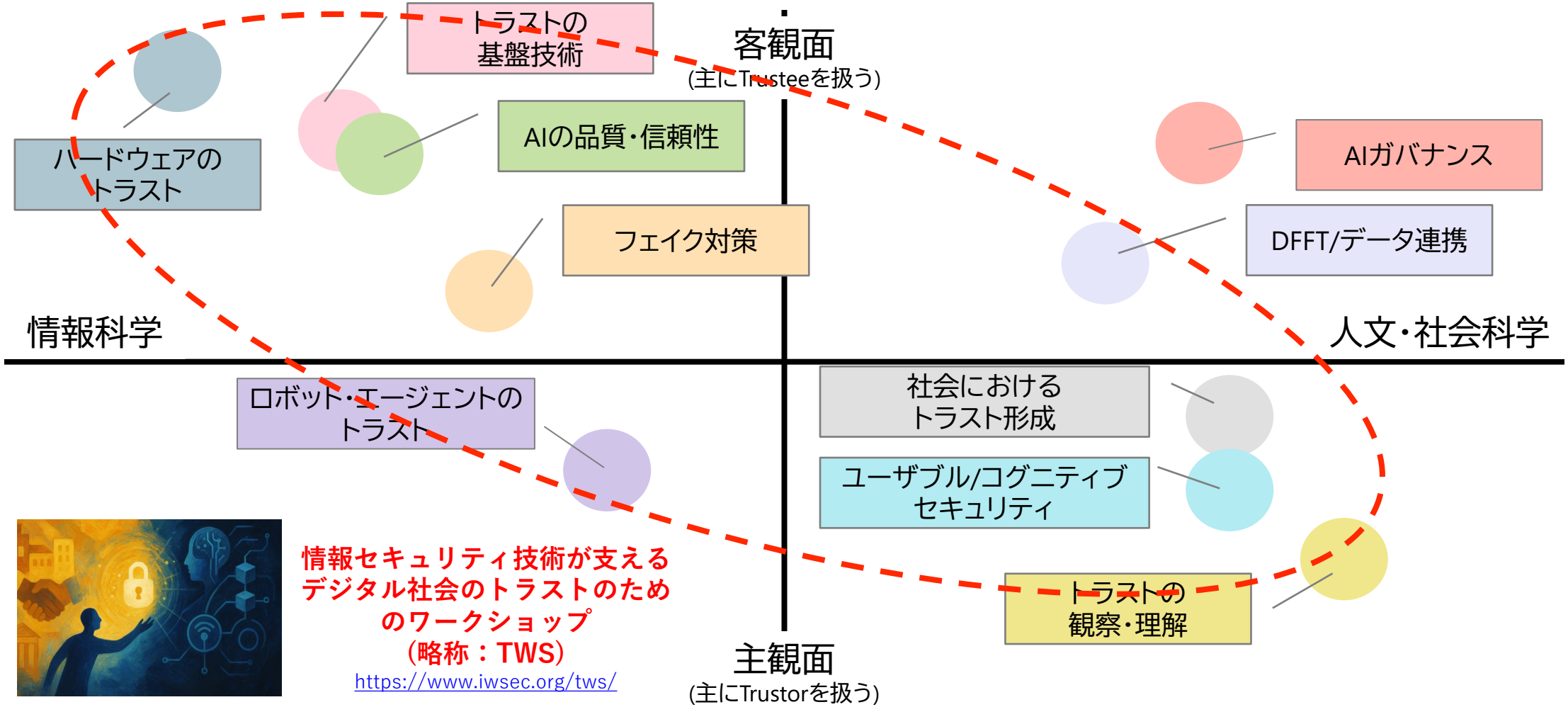
認知科学とセキュリティの
融合分野としての
ユーザブル/コグニティブ
セキュリティなど

情報科学系による客観面からと、人文社会系による主観面からの両面的アプローチが必要になる

トラストワークショップ

情報セキュリティ技術が支えるデジタル社会のトラストのためのワークショップ

客観面と主観面、情報科学と人文・社会科学、それぞれの分野間を
自由に往来できる研究人材の育成と持続的なコミュニティの形成を目的とする



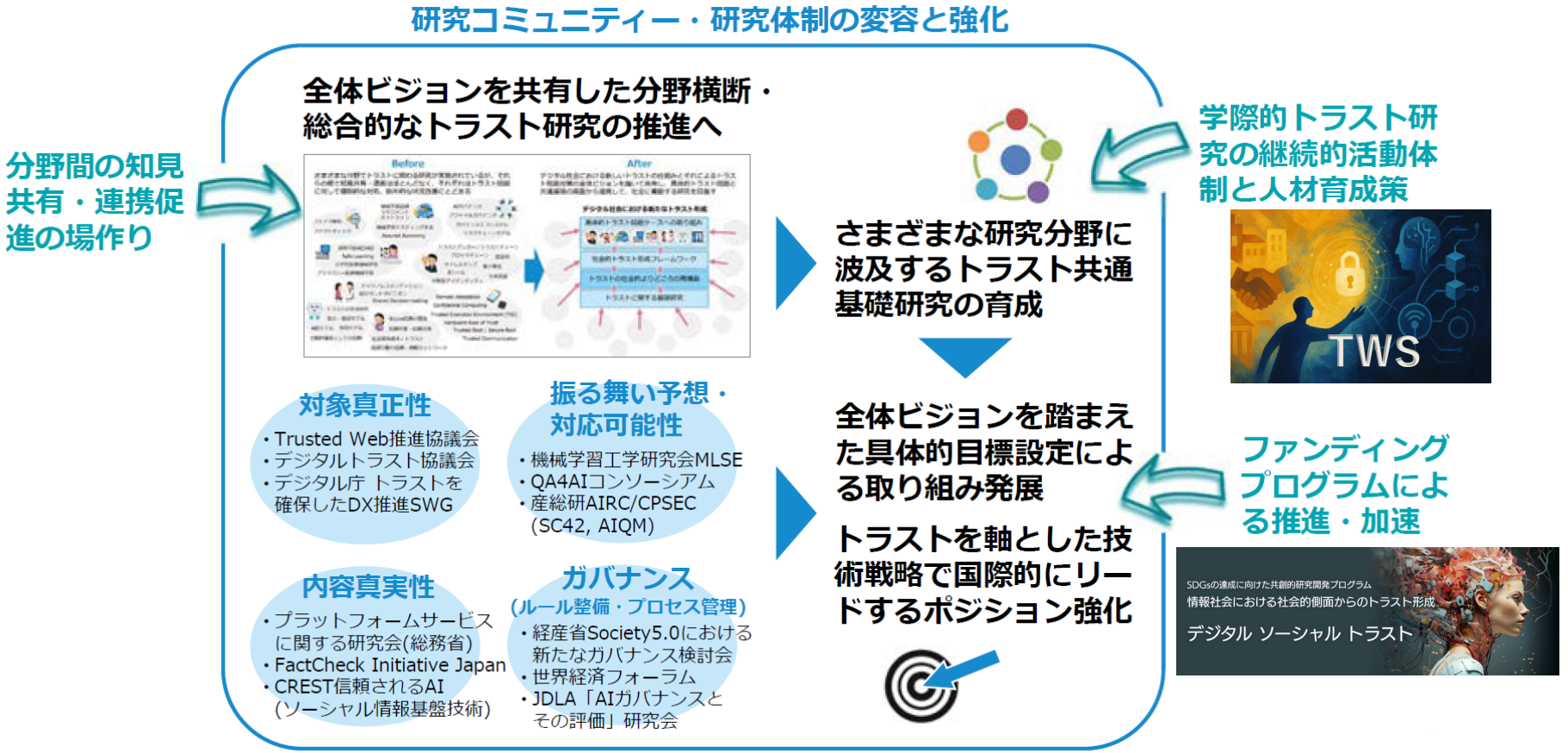
情報セキュリティ技術が支えるデジタル社会のトラストのためのワークショップ（略称：TWS）
https://www.iwsec.org/tws/

図版生成の一部に生成AIを利用しています

小括：分野横断・総合的なトラスト研究の推進

研究コミュニティ・研究体制の変容と強化に向けて

- ファンディングプログラムによる推進・加速
- 学際的トラスト研究の継続的体制と人材育成



戦略プロポーザル：デジタル社会における新たなトラスト形成，CRDS-FY2022-SP-03，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，図4-1，p.27，2022年9月.（演者が一部加筆）

（参考）各分野のトラスト関連動向

ハードウェアのトラスト ハードウェアRoot of Trust、Trusted Execution Environment、リモートアテステーション等でシステム間の信頼関係を人間系の介在なしに検証・確立し、ゼロトラストやコンフィデンシャルコンピューティングへ展開。	**トラストの観察・理解** ABI（Ability, Benevolence, Integrity）モデル、SVS（Salient Value Similarity）モデル、主観確率、安心vs信頼・尺度など、心理・社会学系の尺度やモデルを総合。
トラストの基盤技術 電子署名、タイムスタンプ、eシール、Webサーバ証明書等のトラストサービスによる本人／本物の真正性と改ざん耐性の保証。DID（Decentralized Identity）やVC(Verifiable Credential)、DIW(Digital Identity Wallet)などへの利活用。	**ユーザブル／コグニティブセキュリティ** 人の認知能力等の脆弱性を突く攻撃への対処、使いやすさとセキュリティの両立、認知の傾向を踏まえた設計。ヒューマンファクタ、認知バイアス、セキュリティUX（User eXperience）、フィッシング対策など。
AIの品質・信頼性 Trustworthy AI／AI ELSI（Ethical, Legal and Social Issues）、品質保証・説明可能性・公平性・安全性・プライバシーを総合的に扱う潮流。社会受容面のニーズも強く、活性度が非常に高い。	**ロボット・エージェントのトラスト** ロボット／エージェントと人間の信頼関係・受容・振る舞いの理解（Human Robot Interaction領域）。情報科学と人文社会の横断。非言語行動、信頼較正など。
フェイク対策 ディープフェイクや偽情報に対する検知・検証の技術と運用。社会的な「よりどころ」形成と併せて重要。	**DFFT（Data Free Flow with Trust）／データ連携** 信頼性のある自由なデータ流通を目指す政策・制度・基盤の整備。データガバナンスやID連携など、トラストの基盤技術の応用分野。
AIガバナンス 技術だけでなく制度・ルール・ステークホルダーの関係性を設計し、トラストが社会へ広がる仕組みを整備。	**社会におけるトラスト形成** 技術要素を社会に受容させるための「社会的よりどころ」づくり、制度・ルール設計、分野横断の総合知。

本日のまとめ

- **トラストの本質と「検証可能性」**
- **ハードウェアセキュリティと検証可能性**
- **トラスト研究の多様なアプローチ**

✓ デジタル社会では、アナログ文化に依拠した旧来のトラストから、検証可能なトラストへの転換が必要

✓ 高度化した技術の存在を前提とした、デジタル社会のあるべきトラストを考える必要

✓ フィジカルの証拠をサイバーへ写像し、“継続的に”状態を確かめ続ける世界へ