別　添　４

報告書（案）
別紙３

ＡＩネットワーク化の進展において想定される課題
（ネットワーク化の観点から）

AIネットワーク化の進展において想定される課題（ネットワーク化の観点から）

◆ AIネットワーク化の特徴として、「様々な事業者間における多種多様なAIの接続（複雑な連携）」、「様々な事業者の多種多様なAIを取りまとめる基盤的なAIの構築（情報や権限の集中）」、「様々な事業者間における多種多様なAIによる情報の連携（情報の共有・拡散）」が挙げられる。

このような特徴を活かして、新しいサービスの開発・提供、最適化や効率化、コスト削減などが可能となる一方で、ネットワークという観点に着目した場合、次のような課題（AIネットワーク化の健全な進展を阻害し得る要因）に留意すべきである。

AIネットワーク化の特徴

- 様々な事業者間における多種多様なＡＩの接続（複雑な連携）
- 様々な事業者の多種多様なＡＩを取りまとめる基盤的なＡＩの構築（情報や権限の集中）
- 様々な事業者間における多種多様なＡＩによる情報の連携（情報の共有・拡散）

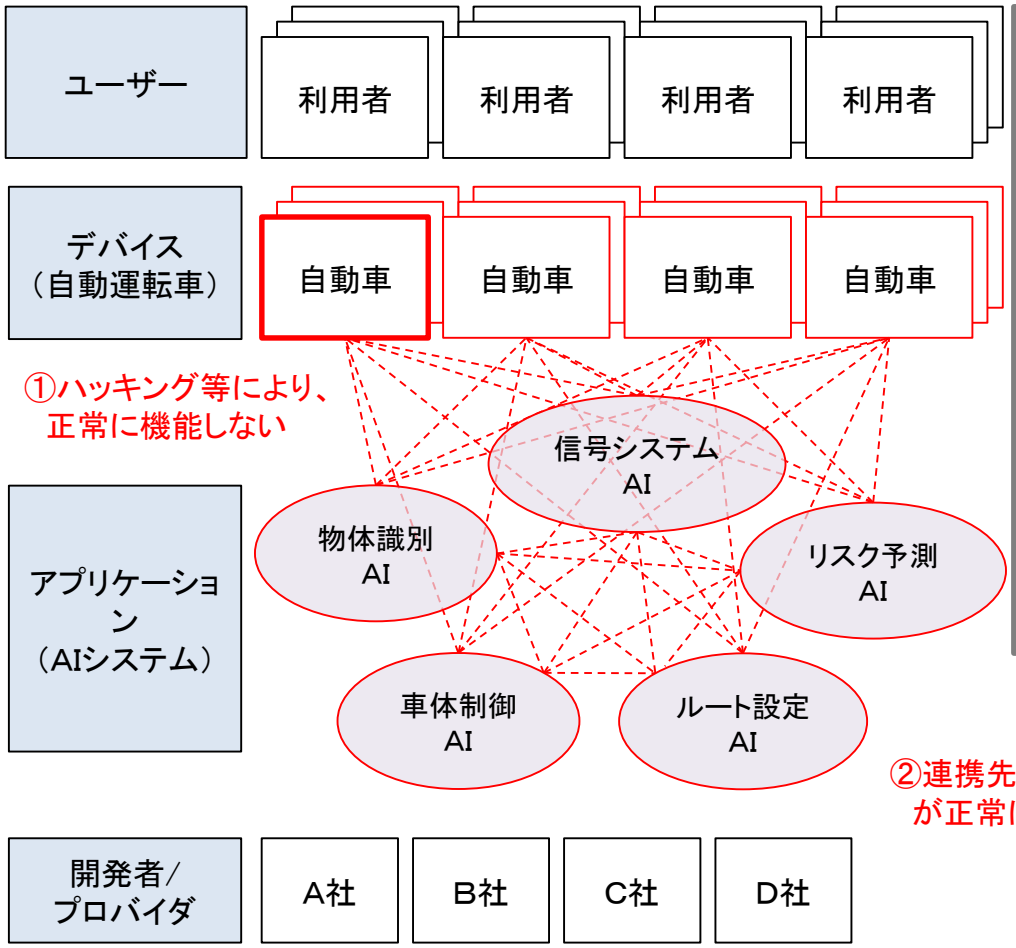
ネットワーク化の観点から想定される課題（例）

- 個別の事業者のトラブル等がシステム全体に波及するおそれ
- ＡＩシステム間の連携・調整が成立しないなどのおそれ
- ＡＩの判断・意思決定を検証できないおそれ（ネットワーク化により、システム全体としてブラックボックス化するおそれ）
- 少数のＡＩの影響力が強くなりすぎるなどのおそれ（少数のＡＩの判断によって企業や個人が不利な立場になるなどのおそれ）
- 領域横断での情報の共有と特定の基盤的なＡＩへの情報の集中によるプライバシー侵害のおそれ
- ＡＩが想定外の動作を行うなどのおそれ

（注）主にＡＩネットワーク化の進展段階２（複数のＡＩシステム相互間のネットワークが形成され、ネットワーク上のＡＩシステムが相互に連携して協調する段階）を想定している。また、ここで掲げる課題については、ＡＩシステムが学習等により自らの出力やプログラムを変化させたり、自律性を有することなどが想定されることを踏まえ、従来のＩＣＴシステムのネットワーク化と比べて、ＡＩネットワーク化によりリスクが新たに生じたり増幅する可能性があるため、ネットワーク化の観点から想定される課題として整理している。

個別の事業者のトラブル等がシステム全体に波及するおそれ

◆ AIシステムが相互に連携して、協調・調整することになるため、例えば、ある事業者において、ハッキングやシステム障害等により一部のAIシステムが正常に機能しなくなった場合や災害等により通信が遮断された場合、あるいは、ある一部のAIシステムの判断・予測等に誤りが生じた場合や虚偽又は不適切なデータが流通した場合などに、他のAIシステムに次々に影響が波及し、ネットワーク全体が正常に機能しなくなったり、期待された効果（適切な判断・予測、マッチング等）が得られなくなったりするおそれがある。

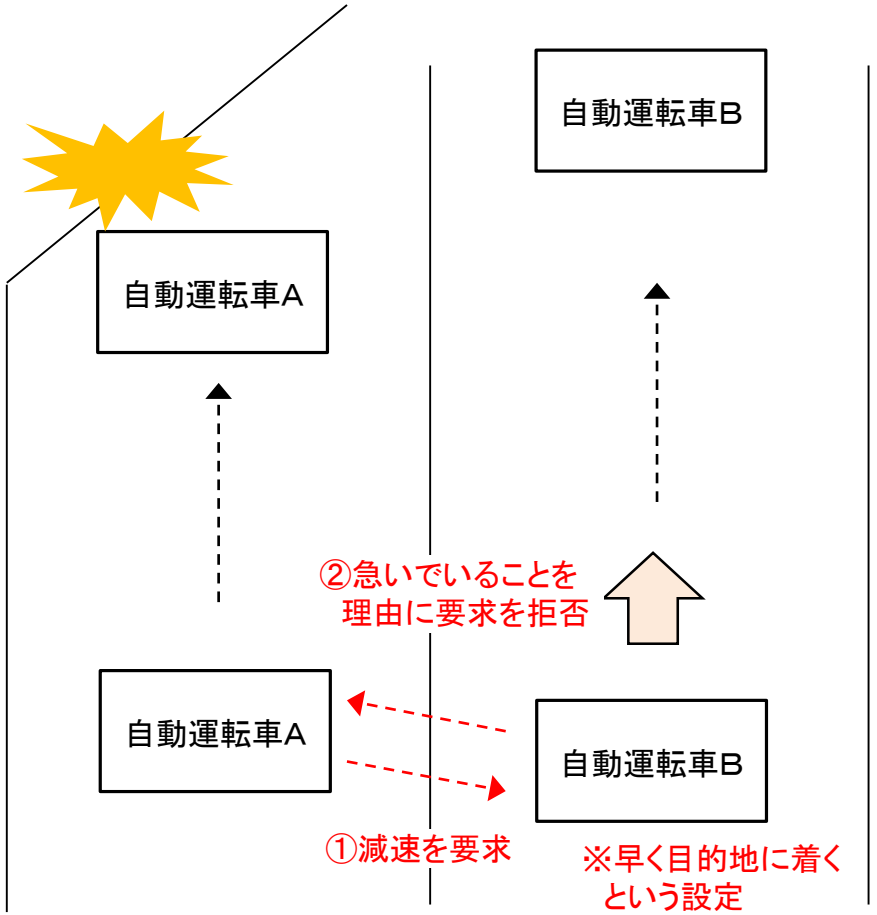


【個別の事業者のトラブル等がシステム全体に波及するおそれ（例）】

- ○ セキュリティ対策等が不十分な事業者が提供しているAIシステムを活用した自動運転車がハッキングされた場合、連携している別のAIシステムが誤作動又は機能不全に陥り、交通事故や交通障害等が発生するおそれがある。*[連携、セキュリティ、安全]*
- ○ サプライチェーン全体として最適化を調整している場合、ある事業者のAIシステムが生産量等を誤って予測した場合や虚偽のデータを流通させた場合、全体としての最適化が実現できない（在庫不足／過剰在庫、原材料不足等）おそれがある。*[連携、データ]*
- ○ 被災地の緊急救命において、救急車や医療機関等の間でリソース配分の最適化の調整を行う場合、ある医療機関のAIシステムが通信の遮断により機能不全に陥ると、適切な情報共有、リソース配分が実現されず、適切な救命活動ができなくなるおそれがある。*[連携、安全]*

AIシステム間の連携・調整が成立しないなどのおそれ

◆ AIシステム間で交渉を行い、連携・調整するケースにおいて、それぞれのAIシステムが自らの目的の達成を優先して調整が成立しなかったり、適切な交渉の権限を有していないことにより取引が成立しなかったりするおそれがある。また、交渉相手のAIシステムが交渉の機能を有していなかったり、交渉に関するプロトコル等が異なったりする場合にも、AIシステム間の連携・調整が成立しない。さらに、AIシステム間の連携・調整が成立したとしても、社会的に望ましい結果にならない（デジタル・カルテル等）おそれがある。

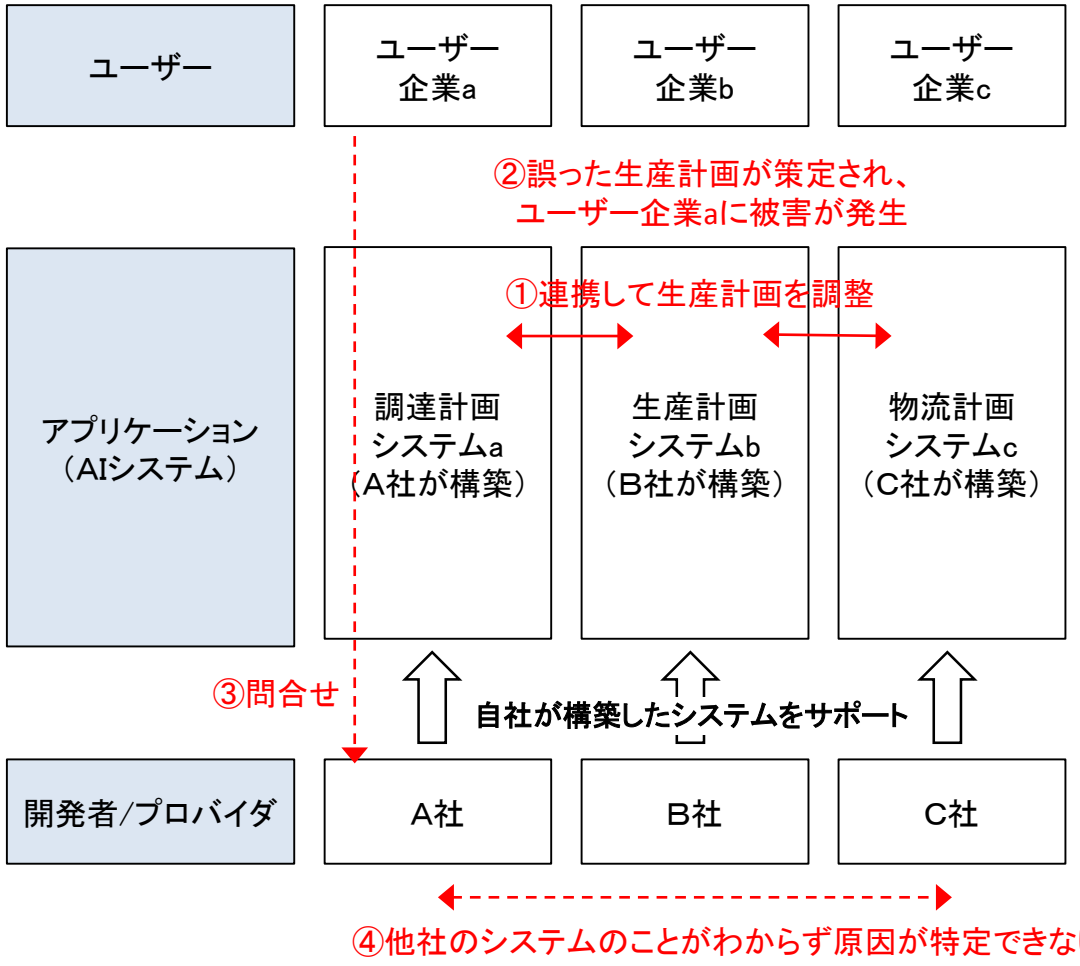


【AIシステム間の連携・調整が成立しないなどのおそれ（例）】

- ○ 自動運転車が車線変更する場合において、A社の自動運転車が車線変更先を走行しているB社の自動運転車に減速するよう要求しても、B社の自動運転車は「なるべく早く目的地に着く」という設定をされていると、減速の要求に応じないというケースが想定される。さらに、車線が減少する場面においては、事故につながるおそれがある。*［連携、安全］*
- ○ AIシステム間で売買契約等における価格交渉を行う場合において、交渉相手のAIシステムが、交渉の権限を有していないにもかかわらず、権限を有しているかのように振る舞うことにより、適切な取引が成立せず、不測の損害を被るケースが想定される。*［連携］*
- ○ 競争関係にある事業者が、それぞれ提供するサービスの価格設定に関し交渉を行う場合において、共通するアルゴリズムを使用していると、協調的価格設定（デジタル・カルテル）を通じて競争が制限されるなどのケースが想定される。*［連携］*

AIの判断・意思決定を検証できないおそれ

◆ AIシステムが相互に連携して、協調・調整することになるため、複数のAIシステムが連携して判断、意思決定を行う場面が想定される。このような場合、複雑な意思決定プロセスになる可能性が高く、また、意思決定プロセス自体がブラックボックス化して、AIシステムが行った判断・意思決定の正誤、精度の検証、トラブル時の原因究明や被害者の救済が困難になるおそれがある。



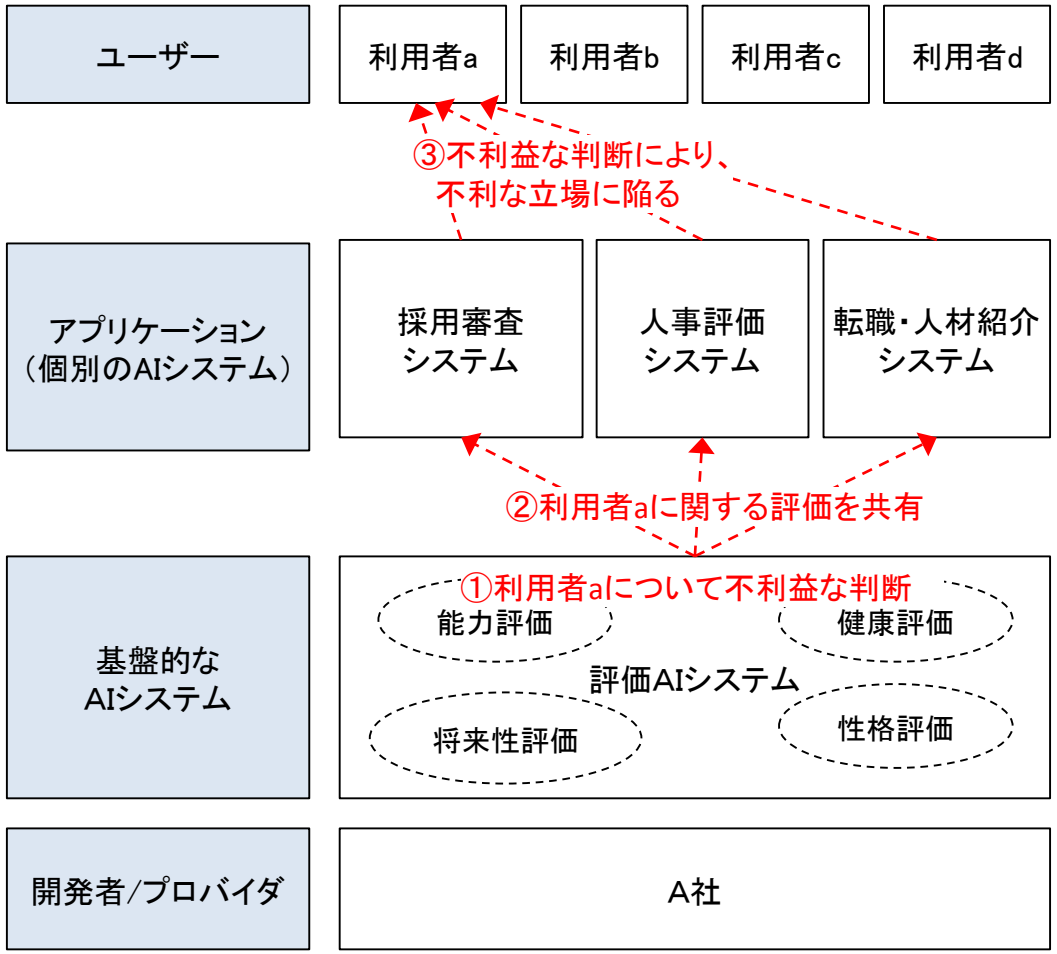
【AIの判断・意思決定を検証できないおそれ（例）】

- ○ サプライチェーン全体として最適化に向けた調整を行っているケースにおいて、個々のAIシステムは、それぞれ最適化を目指して調整するものの、何らかの要因で全体最適とならなかった場合、他社のAIシステムのことがわからず、原因を検証・特定することができないおそれがある。*[連携、ブラックボックス化]*
- ○ 様々な情報（顧客情報、市場動向等）をもとにAIシステムが経営判断を支援しているケースにおいて、それぞれの情報が他の情報の前提となっているような場合、調整自体が複雑となり、人間では、そのプロセスを把握することが困難になるおそれがある。*[連携、ブラックボックス化]*
- ○ 複数のAIシステムが連携してサービスを提供しているケースにおいて、事故等により損害が発生した場合、被害者が、どのAIシステムにより損害がもたらされたのか立証することができず、適切な救済を受けることができないおそれがある。*[連携、ブラックボックス化、安全]*

少数のAIの影響力が強くなりすぎるなどおそれ

◆ AIネットワークのエコシステムにおいては、データに関するネットワーク効果が働きやすいと考えられる。このため、ある特定のAIシステムにデータ等が集中することとなり、その結果として、基盤的なAIシステムが存在する構造が想定される。

◆ このような構造において、例えば、「ヒト」の評価や採用のベースとなる基盤的なAIシステムが独占的・寡占的に提供された場合に、基盤的なAIシステムの判断によって、企業や個人が不利な立場になるおそれがある。また、基盤的なAIシステムの開発者が、基盤的なAIシステムとつながっている様々なアプリケーション（個別のAIシステム）の判断を意図的にコントロールし、特定の企業や個人を不利な立場に陥れることができる可能性がある。



【少数のAIシステムの影響力が強くなりすぎるおそれ（例）】

○ 採用（就職）時に不合格となった場合に、他社の採用においても不合格が続いたり、別の会社に就職しても昇格・昇給が遅れたり、転職時にも良い評価が得られないなど不利な立場に陥るおそれがある。*[連携、正当性・公正性]*

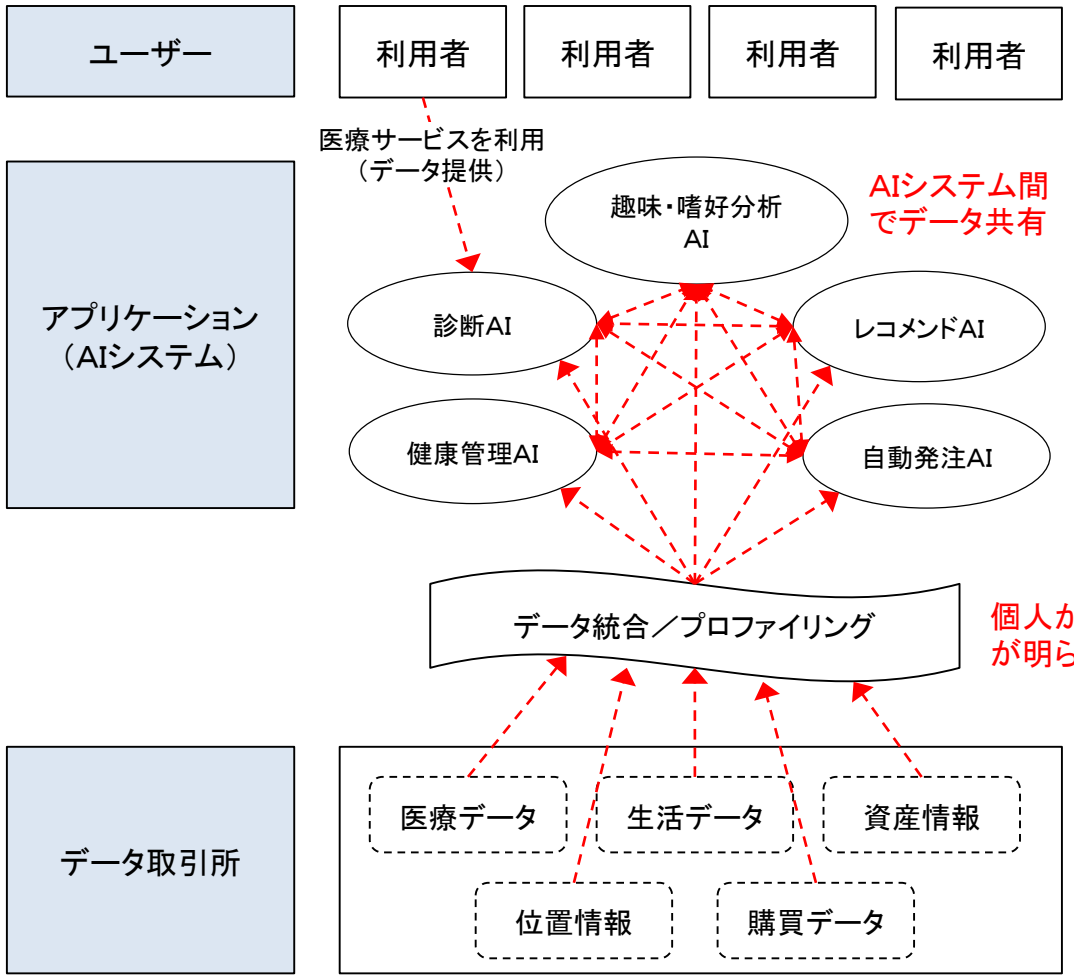
○ 融資の申込みを断られた場合に、どの金融機関等からも融資が受けられずに事業の存続が困難に陥るおそれがある。*[連携、正当性・公正性]*

○ 入国審査で、誤ってテロリストや犯罪者と認定されてしまった場合に、他国にも入国できなかったり、差別的な扱いを受けるおそれがある。*[連携、正当性・公平性]*

プライバシー侵害のおそれ

◆ AIネットワークのエコシステムにおいて、領域横断的なデータ取引所の構築、大量のデータを保有し他者に提供するデータブローカー、様々なデータの統合／プロファイリングを行う事業者の登場が想定される。

◆ このような状況において、意図しない形で、かつ、本人の知らないところでプライバシー性の高い情報が拡散するおそれがあるほか、様々な情報の統合／プロファイリング等により、個人が特定されプライバシーが侵害されるおそれがある。また、あるAIシステムがハッキング等された場合に、ネットワークを通じて他のAIシステムにおいて情報が共有される（流出する）おそれがある。



【プライバシー侵害のおそれ（例）】

○ AIシステム間の調整により、本人同意なく、自動的に個人情報、パーソナルデータが流通するおそれがある。*[連携、プライバシー]*

○ データ取引所等から集められた様々なパーソナルデータが統合／プロファイリングされることにより、個人が特定され、個人に紐付いた行動履歴等が明らかになるおそれがある。*[連携、プライバシー]*

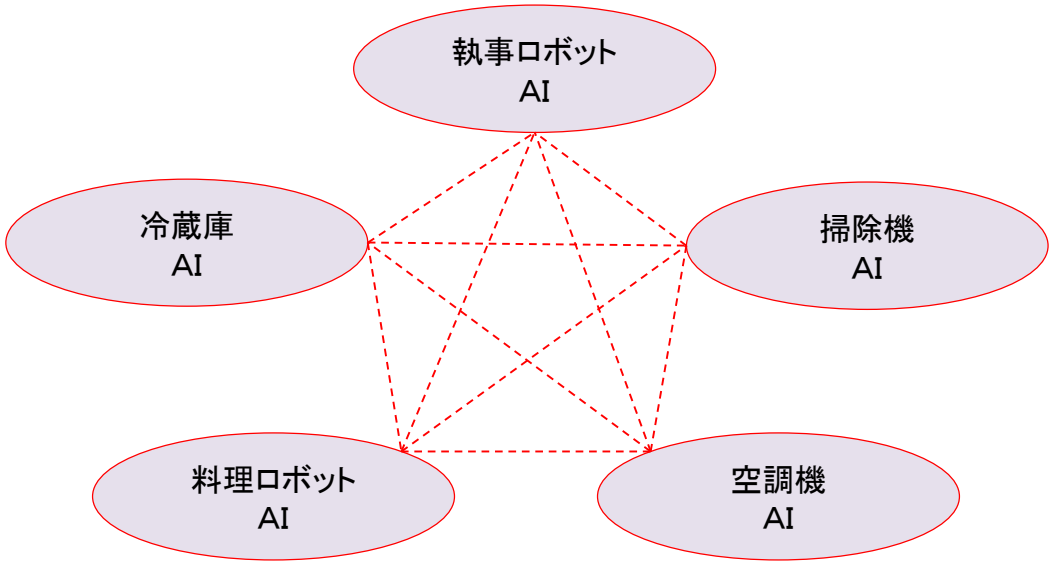
○ あるAIシステムがハッキング等された場合、ネットワークを通じて即座に他のAIシステムに情報が共有され、もともと保有している情報と統合することにより、個人が特定されるおそれがある。*[連携、プライバシー、セキュリティ]*

AIが想定外の動作を行うなどのおそれ

◆ AIシステムがネットワークを形成し、複数のAIシステムが相互に連携して機能することが想定される。このような構造において、どのAIシステムがどのような機能を果たしているのか把握することが困難になるおそれがある。また、AIシステムがネットワーク化されることにより想定外の動作を行うおそれがあるとともに、想定外の不具合が生ずるおそれがある。

◆ AIシステムがネットワーク化されることにより、同じデータが共有されることで、AIシステムの判断の傾向が画一的になるおそれがある。他方で、AIシステムの判断の前提となっているよう条件が成立しなくなった場合、ネットワークにつながっている全てのAIシステムが影響を受け、適切な判断を行うことができなくなるおそれがある。

温度の管理は、
執事ロボットAIの判断？or 空調機AIの判断？



食材の自動注文は、
冷蔵庫AIの判断？or 料理ロボットAIの判断？or 執事ロボットAIの判断？

⇨ それぞれのAIシステムが異なる判断をしたら？？

【AIが想定外の動作を行うなどのおそれ（例）】

○ 家庭内の執事ロボットや空調機などの家電が連携して快適な住環境を確保（例えば、室温の管理）する場合、全体を統制する執事ロボットの判断なのか、個々の家電（空調機）の判断なのかがわからず、仮にそれぞれの判断が相反すると不具合が生ずるおそれがある。*[連携、ブラックボックス化、受容性]*

○ AIシステムを用いて株式や外貨取引等を行う場合、極めて短時間で様々なAIシステム間で交渉・調整することが想定され、どのような交渉・調整が行われたのか人間が把握することが困難になるおそれがある。*[連携、ブラックボックス化、受容性]*

○ 経済予測を行う場合や店舗の来客数の予測を行う場合に、前提となるデータ（GDP速報や金利、天気や通行量など）が共有されることで、AIシステムの判断の傾向が画一的になるおそれがある。*[連携、データ]*