【講演レポート】

AI ガバナンス・セキュリティと政策動向

AI リスク　「技術課題」から「経営の意思決定課題」へ転換

マサチューセッツ工科大学サイバー・AI セキュリティマネジメントコンソーシアム（CAMS）
副統括責任者
慶應義塾大学　特任教授
ミュンヘン工科大学 AI 倫理研究所客員教授　藤末　健三氏

世界の AI パワー比較

現在 AI は米国と中国が覇権国家となっています。以下、サーバー市場、投資、特許、人材別に各国の動向を比較し紹介します。（図１）

(1) AI サーバー市場

一般的なサーバー市場については北米が 4 割、アジア圏が 3 割、EU が 2 割という状況ですが、AI サーバーに絞ると北米が 74%を占有しています。中国は米国からの輸出規制により 5 年前の 25%から現在は 15%に低下しています。ただし、世界地図上でサーバー市場を分析すると、米・中国間で AI サーバーを世界各地にどれだけ作るかの競争が行われていることがオックスフォードの研究機関の調査で把握できます。

(2) AI 研究開発投資額

AI 研究開発投資額を比較すると、米国が 4,700 億ドル、中国が主に国家主導で 1,190 億ドルを投じているのに対し、日本は 60 億ドルにとどまっています。AI について議論する際には、米国や中国が AI の研究開発に莫大な資金を投入している、ということを留意しておかなければなりません。

こうした状況の中、日本は AI ビジネスやスタートアップの議論において海外動向を把握していなかったり、AI のビジネス化やセキュリティマネジメントへの取り組みが遅れてしまうのではないか、と懸念しており、よりグローバルに、より幅広く動向を把握していく必要があると思っています。

(3) AI 関連特許

AI 特許シェアについては中国が約 5 割とトップを占め、米国 14%、日本 11%となっています。ただし、中国は国内特許がほとんどのため、特許比較はあまり重視しておらず、先に紹介した AI 研究開発投資額の比較の方を重視しています。

(4) AI 研究者数・トップ人材シェア

AI 研究者数を比較しても米国が圧倒的に多く、次に中国と続きますが、中国生まれの研究者が米国から中国に戻ったり、新たな AI アルゴリズムが開発されるなど、今後は中国の AI 研究者が多くを占めていくだろうと予想しています。また、インドも急激な勢いで研究者の育成に取り組んでいます。

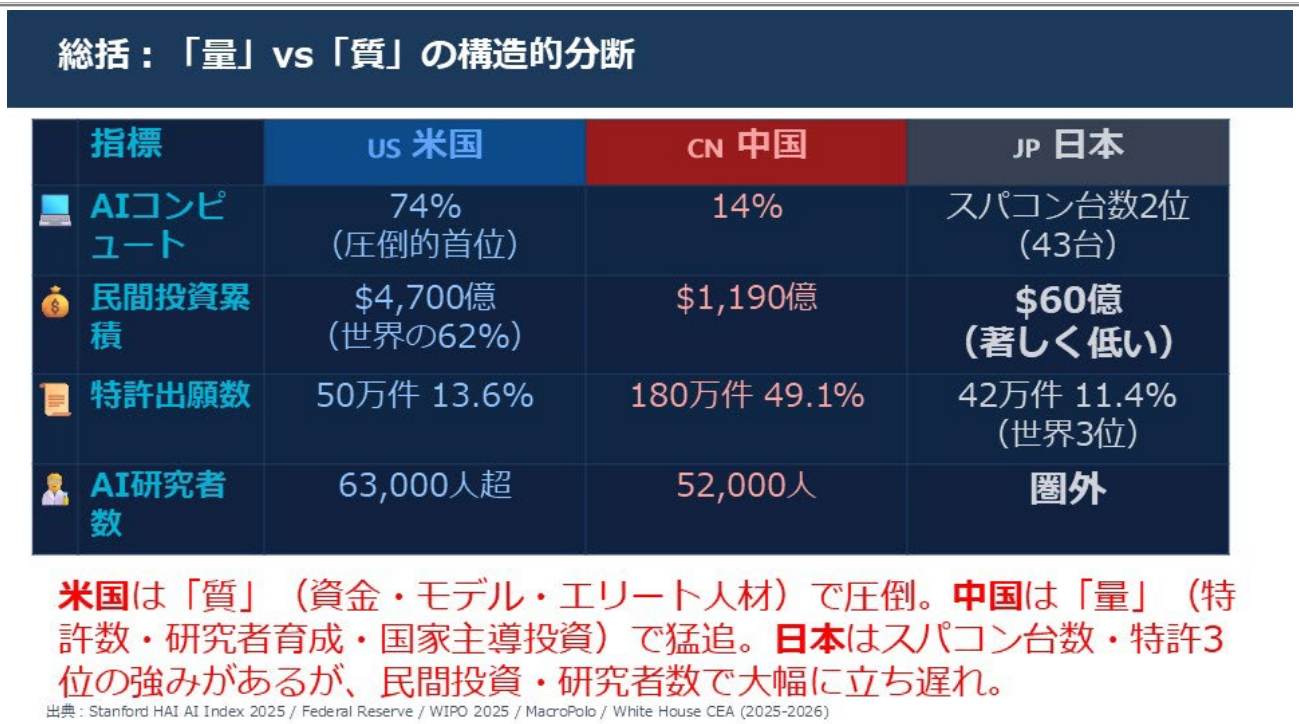

図 1.　世界の AI パワー比較

AI 覇権の民主化：国際社会への提言

(1) 提言

トルコで開催された GLOBAL AI GOVERNANCE SUMMIT では、米国や中国が AI 覇権を握る状況下において、以下の 3 点が提言されました。（図 2）

- サーバーの公共化（公共化としての AI）
- アルゴリズムのオープンソース化
- AI データのコモン化（共有化）

AI データのコモン化の例として、インドでは患者データが安価で買われています。AI で重要とされる正確なデータ、特にバイオデータが資本で集められてしまうと、お金のない国民が AI 診断を受けられないような状況になりかねません。そこで AI のデータ共有化により世界中の人たちが AI の恩恵を受けられるようにしよう、と提案されました。

今後、日本でも同様の議論がされていくと思われますが、AI サービスが普及していく中、潤沢な資金を持つ企業や個人が AI を独占するような時代の到来に対し、どう対処していくか、がこれから求められることだと思います。

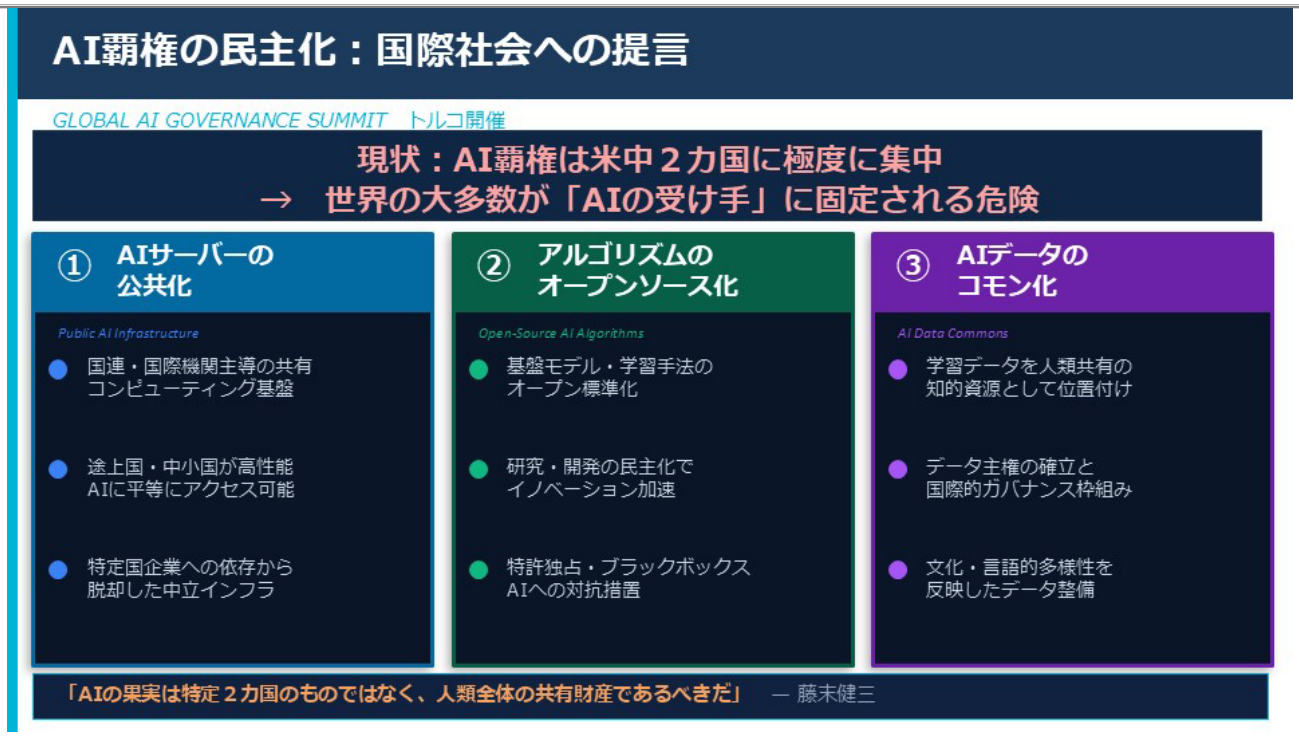

図 2.　AI 覇権の民主化：国際社会への提言

（2）法規制

EU では人権保護やプライバシー保護をメインとした EU AI Act が成立して 2 年が経ちました。日本でもこの法律に準じたルールを作ろうとしています。一方米国では、Google や Amazon などで AI の判断結果による人種差別や性差別などのトラブルが発生しており、法規制よりも民間企業によるガイドラインの策定等で規制が行われています。

もう一つ、議論が必要視されているのが兵器としての AI 規制です。AI を利用した兵器をどう規制するか国連で議論されていますが、なかなか進まない状況です。

AI 規制について、EU は人権を侵害せずにどう活用するか、米国は産業・ビジネス成長に、中国は国家の安全保障・管理に AI をどう活用するか、とそれぞれ求める方向性が異なるため、世界で統合された AI 管理の議論はなかなか進まないのではないか、と感じています。

AI 時代のセキュリティ

（1）AI そのものを守る

AI 時代のセキュリティは「AI で守る」段階から「AI そのものを守る」段階へと移行してきています。

AI に汚染されたデータを学習させることで判断を狂わせたり、アルゴリズムへの細工、AI のフィードバックループを狂わせる、誤った議論に誘導してしまうなどがあります。AI の回転スピードが速すぎるために、マルウェアや悪い情報を埋め込めばどんどん悪い方に進んでしまうのが AI の特徴だと思います。（図 3）

図 3.　AI 時代のセキュリティ：新たな攻撃面（Attack Surface）

（2）セキュリティ知識の共有

AI セキュリティもサイバーセキュリティと同様ですが、日本はテクノロジー側に寄りがちですが、目に見える領域だけでなく、根幹にある文化、社員のセキュリティに対する意識や、経営者がセキュリティを理解し、トラブル発生時に即座に対応、優先順位を的確に判断でき、社員が主体的に参加することが重要であり、その前提としてセキュリティに対する知識を共有することが重要となります。（図 4）

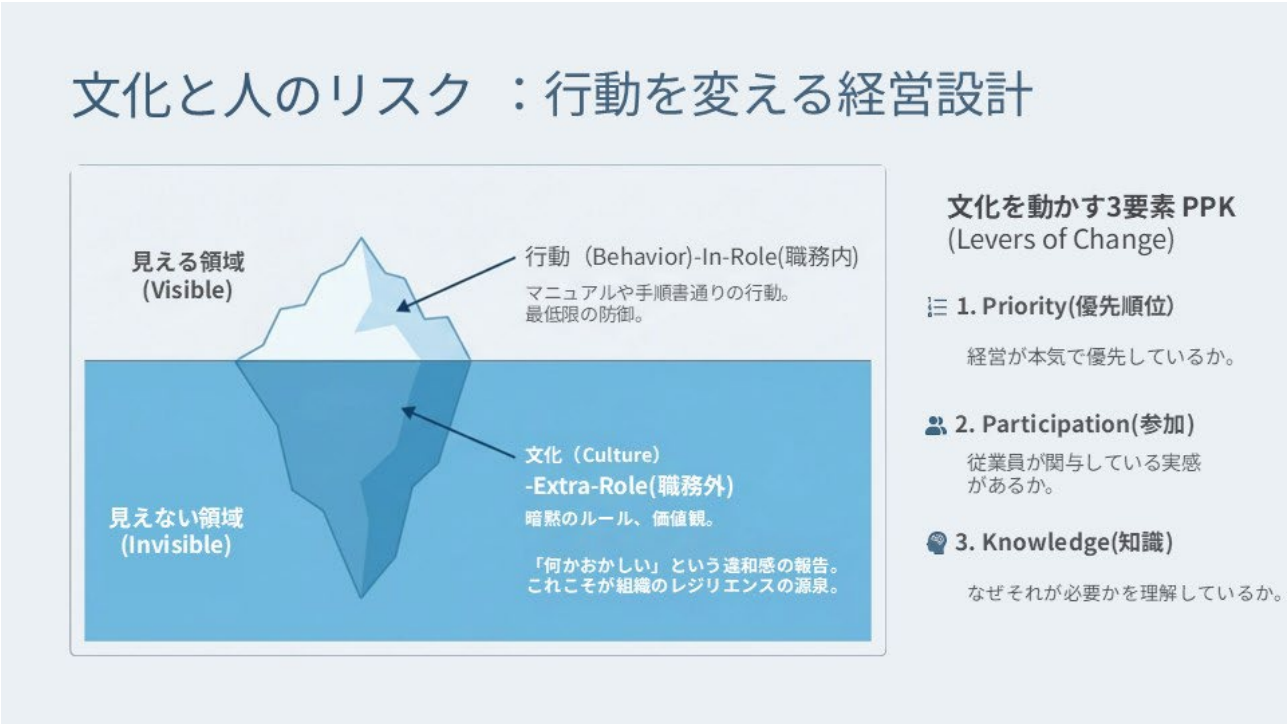

図 4.　文化と人のリスク：行動を変える経営設計

（3）成熟度モデルとリスクの定量化

これまでのセキュリティは「問題を起こしてはいけない」「当事者の責任」でしたが、昨今では AI、サイバーともに攻撃や問題発生が当たり前になっており、インシデント発生後いかにレジリエンスを高めて復活させるか、原因究明するか、などの評価基準が成熟度モデルにより変わります。

また、リスクを定量化し、技術指標そのままではなく、損失や費用がいくらになるか、といった経営指標に置き換えることが重要です。

（4）サイバーAI レジリエンス：経営の設計（5 つの決断）

経営者には、インシデント発生後、①優先順位、②停止権限、③代替手段、④復旧順序、⑤外部連携の判断が求められます。今後インシデントが発生した際に、経営者がどう対応するかが重要になります。（図 5）

図 5. サイバーAI レジリエンス：経営の設計（五つの決断）

（5）AI セキュリティの課題

「信頼の 72 時間」とよく言われていますが、インシデント発生後どれだけ早く対応できるかがサイバーセキュリティ、AI セキュリティの課題であると考えます。（図 6）

経営者は「止まらない七つの原則」として、以下を設計しておくことが重要です。

- 原則 1：事業継続を最上位目的に置く
- 原則 2：例外を資産負債として管理する
- 原則 3：代替手段を設計しておく
- 原則 4：意思決定の RACI を決める
- 原則 5：更新型コミュニケーションにする
- 原則 6：第三者リスクを前提にする
- 原則 7：事故後に学習する

図 6. 危機時コミュニケーション：信頼の 72 時間

ISO/IEC 42001（AIMS）「説明できる AI」を実装する

（1）AIMS とは

AI マネジメントシステム（AIMS）は、AI を情報セキュリティマネジメントシステム（ISMS）同様の管理対象として PDCA サイクルを回しながら継続改善していくものです。

- AI を品質（Q）や情報セキュリティ（IS）と同じ管理対象に
- PDCA サイクルで継続的に改善
- Clause 4～10：範囲→責任→計画→支援→運用→評価→改善
- Annex A：9 領域・38 の参照コントロール
- SoA（適用宣言書）で説明責任を担保

「AIMS100 日ロードマップ」（図 7）のように、全体を把握→システムの管理体制を構築→確認→改善する仕組みをどう作るかがポイントとなります。

ここでのポイントは、ISMS と AIMS について、管理を分断（サイロ化）させず、既存のリスク管理と統合する必要がある点です。後ほどご紹介する MIT の CAMS においても、AI セキュリティとサイバーセキュリティの一体化について議論が始まっています。

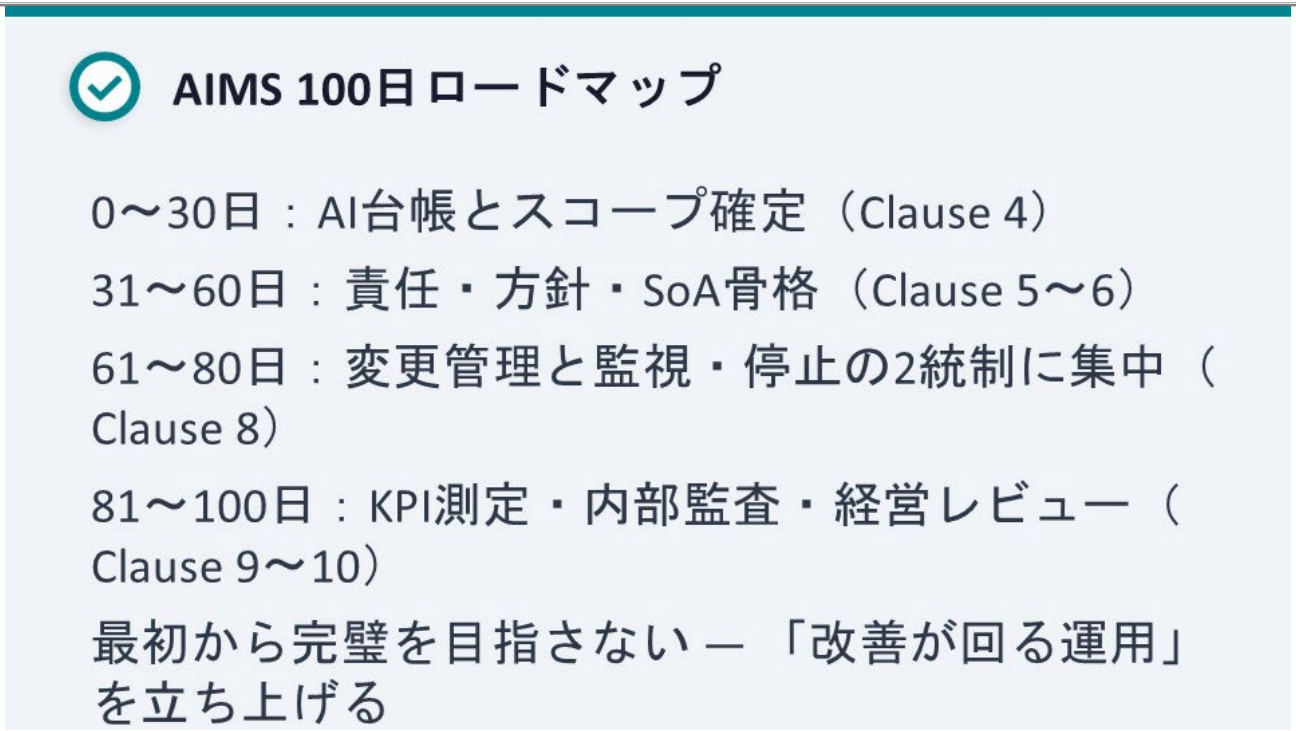

図 7. AIMS100 日ロードマップ

（2）規制の洪水を「設計」に変える

企業はさまざま規制をいかにきれいにデザインしてまとめていくかが重要ですが、特に重要なのが「三線モデル」という、現場・管理部門・監査の各データを三つのラインで整理し、取締役会では 1 ページの資料で報告できるように設計することです。複雑なマネジメントを単純化し、かつシンプルにすることが大きな流れとして必要になると言われています。

（3）サイバーAI の文化

サイバーAI の文化として、経営者や従業員の理解、知識の共有が重要であり、自社に限らず、サプライチェーン間で共通した「セキュリティを守る文化」を醸成することが必要になります。

経営には、優先順位をどこに置くかを決め、皆が参加し、知識を共有することが重要で、そのために職務内の安全行動をマニュアル化したり、おかしいと思ったら声を上げる、といった体制を作っていくことが重要です。

CAMS について

CAMS はマサチューセッツ工科大学（MIT）の経営大学院が行っているプロジェクトで、戦略・ガバナンス・マネジメント・組織の四つの柱をテーマに「経営」がどうあるべきかを研究・発信しています。

日本の企業に対しても経営手法やインシデント発生を想定した体制作りや、ボードメンバーがどうすべきかなどを示唆しています。

CAMS にはさまざま分野の研究者が携わっており、最新のデータや技術をいかに最新の経営に反映させるかまで見ることができ、それが強みでもあります。また、さまざまなインシデントデータも収集しています。これまでの日本語の壁によりサイバーが守られている時代は終わり、外国からの攻撃も増えてくると予想していますので、日本でも役に立つと思っています。

CAMS の議論において、どういう結果をボードメンバーに提示し、どう判断させなければならないかが非常に重要になります。

したがって、AIMS の土台にマネジメントやガバナンスを置いていくことが CAMS や MIT の役割だと考えています。

最後に

AI セキュリティ、サイバーセキュリティの課題は経営の設計問題であると考えます。攻撃を受けることを前提に、いかに攻撃で止められる確率を下げるか、止まった場合の復活、レジリエンス強靭化といいますが、レジリエンスをどう作っていくか、攻撃や問題発生後に学習する、といったことをどう設計するかが必要になります。

AIMS の普及によるプラットフォームの確立後、国際動向などを参考にしつつ、さまざまなコンサルティングや社員教育などを付加していくようになるのではないか、と思います。

藤末　健三氏

マサチューセッツ工科大学サイバー・ＡＩセキュリティマネジメントコンソーシアム（ＣＡＭＳ）　副統括責任者／慶應義塾大学　特任教授／ミュンヘン工科大学 AI 倫理研究所　客員教授

MIT ビジネススクールのサイバー・ＡＩセキュリティマネジメントコンソーシアム（CAMS）アソシエイト・ディレクター(副統括責任者)。

ミュンヘン工科大学ＡＩ倫理研究所客員教授

慶應義塾大学 特任教授、インド工科大学ハイデラバード校特任教授、韓国先端科学技術院特任教授、オックスフォード大学インターネットインスチチュート前上級客員研究員など、国際的な研究ネットワークを構築。

MIT スローン経営大学院およびハーバード・ケネディスクールで修士学位を取得。早稲田大学および東京工業大学で博士号を取得。

学術・政治活動に加え、プロボクサーライセンスを保持し、現在もトレーニングを継続

本内容は、2026 年 3 月 30 日から 4 月 20 日にかけてオンデマンド配信された JIPDEC セミナー「AI のリスクマネジメントと AI マネジメントシステム（AIMS）認証の最新動向」での講演内容を取りまとめたものです。