THE TRILATERAL COMMISSION

公益財団法人 国際文化会館
International House of Japan

＜別紙: 三極委員会「AI、ヒューマンコントロール、戦略的安定性に関するスタディグループ」について＞

概要

本スタディグループは、人工知能（AI）が安全保障、とりわけ核抑止をはじめとする戦略的安定性に及ぼす影響を、三極委員会の三地域(アジア太平洋・北米・欧州)が知見を持ち寄って検討する取り組みです。武力行使、エスカレーション、戦略的安定性に関わる意思決定において、意味のあるヒューマンコントロール、説明責任、責任をいかに維持するかという、AI 時代の中核的な課題に取り組みます。

背景と問題意識

AI は、人間の判断、産業社会、政治権力の関係を変えつつあり、現代を規定する力の一つとなっています。その影響は、安全保障と戦略的安定性において、とりわけ重大です。AI はすでに、情報分析、サイバー防衛、兵站、自律型システム、重要インフラ防護、標的選定支援、意思決定支援、指揮統制といった領域を形づくりつつあります。AI を強力なものにしている特性(速度、規模、パターン認識、適応的な意思決定支援)は、同時に、意思決定の時間を圧縮し、責任の所在を不明確にし、意図せざるエスカレーションのリスクを高めうるものでもあります。責任をもって用いれば AI はレジリエンス、法令遵守、危機管理を強化しうるが、十分な安全策なしに用いれば、これらのリスクが顕在化しかねません。

安全保障・防衛における AI は、既存の枠組みがいまだ十分に対処できていないガバナンスの領域に入りつつあります。自律型致死兵器をめぐる議論は不可欠ですが、問題の一部を捉えているにすぎません。また、民生分野の AI ガバナンスは、安全性、透明性、プライバシー、イノベーションを扱いますが、国家が権力を行使する制度・インフラ・意思決定プロセスに AI が組み込まれていく際の戦略的な帰結は、未解決のまま残されています。こうした問いは、既存の国際的な議論のなかでも十分に掘り下げられておらず、本スタディグループはこの空白に取り組みます。

なぜ三極委員会か

三極委員会は、条約フォーラムでも、軍事同盟でも、技術標準を定める団体でもありません。その価値は、危機によって分断が固定化する前に、政治、ビジネス、技術、法律、学界、安全保障の各コミュニティのリーダーたちが共に考えられる、率直な対話の場を創り出すことにあります。本スタディグループは、次のような三極委員会ならではの強みを活かして検討を進めます。

1. 戦略、技術、法律、産業、政治という異なる視座を結びつけ、インド太平洋と欧州大西洋の両地域をつなぐことができる点。無制限の AI 開発競争と非現実的な凍結との間で、責任ある道筋を描く助けとなりうる、柔軟で軽量な対話のプラットフォームを提供します。
2. 創設以来培ってきた金融分野のネットワークを通じて、AI 安全保障をめぐる経済的側面に、安全保障に特化したフォーラムにはない視点を持てる点。AI をめぐる資本の流れ、政府調達、産業構造の変化といった論点は、AI 能力がどのように拡散し、誰がそれを管理するのかを左右するものであり、三極委員会が歴史的に有してきた強みの核心に位置します。
3. 政権交代を越えて専門家を招集できる非党派性を備えている点。これにより、現在どの政府プロセスでも維持されていない専門知の継続性を保つことができます。
4. 各地域の多様な政策・技術コミュニティとの接点を持ち、異なる政治・社会的文脈を踏まえた対話を重ねられる点。とりわけアジア・太平洋地域は、倫理と人間の責任をめぐる多様な思想的伝統など、既存の議論で十分に扱われていない視座を提供しえます。中国、グローバルサウス、技術産業の専門家との協議も目指します。

検討の焦点

本スタディグループは、安全保障・防衛の応用領域を横断して、AI を備えたシステムが、人間が判断を下し、責任を割り当て、リスクを管理する条件をいかに変えつつあるかを検討します。とりわけ、速度、自律性、不確実性、そして戦略的帰結が交わる領域に注意を払い、次の論点を中心に検討を進めます。

1. **核の指揮統制への AI の関与** 核の指揮統制支援、戦略的早期警戒、エスカレーション評価、意思決定支援といった、最も重大な帰結を持つ領域において、AI がもたらす含意を検討します。ここでの AI は、標的選定の高速化による意思決定時間の圧縮、意思決定支援ツールへの自覚されない依存など、既知のリスクを増幅する要因として作用しえます。AI の使用が許容されない領域、厳格なヒューマンコントロールが不可欠な領域、追加的な安全策が必要な領域を特定します。AI と核兵器に関する規範的な了解は、ほぼ普遍的な懸念が共有される稀な領域として、達成可能であると広く見なされています。

2. **AI とバイオセキュリティの接点** AI と生命科学の交わりは、AI を活用した設計・合成が、非国家主体によるものも含め、危険な病原体の開発への障壁を下げうるという、核と同等に重大なリスクを提起します。これは、三地域すべてが積極的に関与している領域でもあり、生物兵器禁止条約の下での検証取り決めに向けた作業などと連携して検討を進めます。
3. **異なる AI システム間の相互作用** 異なるデータ、価値観、前提の上に構築された AI モデルに基づく部隊や意思決定システムが、危機において互いに対峙したとき何が起こるか、という問いは、既存の国際的な議論では十分に扱われていません。そうしたシステムが機械の速度で相互作用するとき、文化的・戦略的な前提が目に見えないところで乖離し、いずれの側も意図せず理解もしないかたちでエスカレーションが進行しえます。本スタディグループは、異なる AI モデルに依拠する部隊が相互作用する際に適用される、交戦規則に相当する共通の期待、連絡の手順、自制措置を検討します。

これらを貫く視座として、自律型システムのライフサイクル全体を通じて責任がいかに維持されるか、機械の速度という条件の下で諸制度がいかに時間、規律、人間の判断を保つか、そして高次の原則が、オペレーター、政策立案者、技術者、法律家、政治指導者にとっての実践的な指針へといかに翻訳されるかを検討します。

また、本スタディグループは、国連、特定通常兵器使用禁止制限条約（CCW）、REAIM（軍事領域における責任ある AI に関するサミット）、NATO、EU、各国政府などの取り組みを補完し、ハイレベルの原則と実務的な指針との間の橋渡しを目指します。とりわけ、2026 年 10 月の国連総会第一委員会での審議、生物兵器禁止条約（BWC）の作業部会、国連のサイバーセキュリティに関する常設プロセスの動向を踏まえて検討を進めます。

検討の進め方と成果物

本スタディグループは、三地域から選出された共同座長(Co-Chair)を中心に、少人数の専門家が参加する体制で進めます。主たる成果物は、AI、ヒューマンコントロール、戦略的安定性に関する簡潔なイシューペーパーであり、主要なリスク、形成されつつある合意、未解決の問い、そして議論のための実践的な選択肢を整理するものです。2026 年秋以降に開催される三地域の地域会合を通じて予備的な知見を共有し、議論を深めたうえで、2027 年の提言の取りまとめを目指します。
本検討の具体的な体制は、以下を予定しています。

	アジア太平洋	欧州	北米
共同座長 **(Co-Chairs)**	**神保謙** 慶應義塾大学教授／国際文化会館常務理事	**ロベルト・チンゴラーニ(Roberto Cingolani)氏** レオナルド社 前 CEO	**ダグ・ベック(Doug Beck)氏** 新アメリカ安全保障センター理事
アドバイザー **(Advisors)**	**尹炳世（ユン・ビョンセ／Yun Byung-se)氏** ソウル国際法アカデミー(SILA)会長、責任ある AI の軍事利用に関するグローバル委員会（GC REAIM)前委員長、元韓国外交部長官） **アーキッシュ・ミッタル(Archish Mittal)氏** Iron Pillar Fund 特別顧問	**マリーチェ・スハーケ(Marietje Schaake)氏** スタンフォード大学ノンレジデント・フェロー、元欧州議会議員 **ハコボ・ロア・ビセンス(Jacobo Roa Vicens)氏** JP モルガン・チェース 機械学習センター・オブ・エクセレンス、ユニバーシティ・カレッジ・ロンドン	**トム・ジェンキンス(Tom Jenkins)氏** オープンテキスト社 会長 **マイケル・B・グリーンウォルド(Michael B. Greenwald)氏** アマゾン ウェブ サービス(AWS)金融イノベーション・デジタル資産グローバル責任者

（参考）アジア太平洋委員会の役割

アジア太平洋委員会は、既存の国際的議論では十分に反映されてこなかった視点の提供を担います。具体的には、倫理と人間の責任をめぐる東アジアの思想的伝統、軍事と民生の境界が曖昧になる「デュアルユース」から「ミックスドユース」への移行と民生側の倫理のあり方、単一の価値体系で開発された AI に文化的な価値判断を取り戻そうとする研究の動向などを想定しています。