

(仮訳)

2024年5月21日のAIソウル・サミット首脳セッション出席国による
安全、革新的で包摂的なAIのためのソウル宣言

1. 我々、オーストラリア、カナダ、欧州連合、フランス、ドイツ、イタリア、日本、韓国、シンガポール、英国及び米国を代表する世界の首脳は、2024年5月21日のAIソウル・サミットに集い、AIの前例のない進展と我々の経済及び社会への影響に直面し、AIに関する国際協力及び対話の促進に向けて共に取り組むことを確認する。
2. 2023年11月に英国のブレッチリー・パークで開催されたAI安全性サミットにおける作業を踏まえ、我々は、AIの安全性、革新性及び包摂性は相互に関連する目標であること、及び、AIの設計、開発、導入及び利用がもたらす、あるいはもたらす可能性のある幅広い機会と課題に対処するために、AIガバナンスに関する国際的な議論におけるこれらの優先事項を包含することが重要であることを認識する。
3. 我々は、AIの安全、安心で信頼できる設計、開発、導入及び利用を確保するために、便益を最大化し、AIによる広範なリスクに対処するためのリスクベースのアプローチに沿った、AIガバナンス枠組間の相互運用可能性の重要性を認識する。我々は、高度なAIシステムを開発する組織向けの広島プロセス国際行動規範の運用を支援することに引き続き集中する。我々は、最先端AIを開発し導入する組織の特別な責任を認識し、この観点から、最先端AI安全性コミットメントに留意する。
4. 我々は、AIセーフティ・インスティテュート、研究プログラム及び／又は監督機関を含むその他の関連組織を創設又は拡大するための、この宣言への参加国による既存かつ継続中の努力を支持し、我々は、これらの組織間のネットワークを育成することにより安全性研究に関する協力を促進し、ベストプラクティスを共有するよう努める。この観点から、我々は、この宣言に添付する「AI安全性の科学に関する国際協力に向けたソウル意図表明」を歓迎する。
5. 我々は、世界最大の諸課題に対処するために人間中心のAIを活用し、民主主義的価値、法の支配、人権、基本的自由及びプライバシーを保護及び促進し、各国間及び各国内でのAIとデジタル格差を埋め、それにより人間の幸福の前進に寄与し、並びに、国連持続可能な開発目標の前進のためのものを含むAI

の実践的な適用を支援するため、A Iの安全性、革新性及び包摂性を前進させるための国際協力の強化を求める。

6. 我々は、安全、革新的で包摂的なA Iエコシステムを促進する、リスクベースのアプローチを含む政策及びガバナンス枠組を主張する。枠組は、人間の創造性とA Iの開発及び利用との間の好循環を促進し、社会文化的、言語的及びジェンダー多様性を促進し、商業的及び公的に利用可能なA Iシステムのライフサイクルを通じて、技術及びインフラの環境的に持続可能な開発及び利用を促進すべきである。
7. 我々は、安全、革新的で包摂的なA Iエコシステムを育むために、政府、民間セクター、学術界及び市民社会を含むマルチ・ステークホルダーによる積極的な協力の重要性、及び、国境を越え、かつ学際的な協力の重要性を確認する。あらゆる国がA Iの便益及びリスクによる影響を受けることを認識しつつ、我々は、A Iガバナンスを取り巻く議論に幅広い国際的なステークホルダーを積極的に含めていく。
8. 我々は、国連及びその関連機関、G 7、G 2 0、経済協力開発機構（O E C D）、欧州評議会及びGlobal Partnership on AI（G P A I）における、その他の国際的取組への関与を通じて、A Iガバナンスに関する国際協力を強化することを目指す。この観点から、我々は、広島A Iプロセス・フレンズグループを認識し、最近改定されたO E C DのA I原則、並びにA Iシステムのセーフガードの必要性及びA Iを便益のために開発・導入・利用するとの責務に関する国際的な認識を強固にした「持続可能な開発のための安全、安心で信頼できるA Iシステムに係る機会確保」に関する国連総会決議の最近のコンセンサスによる採択を歓迎し、2 0 2 4年9月の未来サミットに先立つグローバル・デジタル・コンパクトに関する議論を歓迎し、国連事務総長のA Iハイレベル諮問機関の最終報告書を期待する。
9. A Iの安全性、革新性及び包摂性を促進するA Iガバナンスに関する議論を前進させるためのハイレベルのフォーラムとしてのA Iサミットにおける対話の価値を認識しつつ、我々は、フランスで開催予定のA Iアクション・サミットにおける三回目の集まりに期待する。

付録：A I安全性の科学に関する国際協力に向けたソウル意図表明

(仮訳)

付録

A I 安全性の科学に関する国際協力に向けたソウル意図表明

1. 2024年5月21日のA I ソウル・サミットに集まり、2023年11月2日のブレッチリー・パークにおけるA I 安全性サミットに続き、ブレッチリー首脳セッションのセッション成果に関する安全性評価に係る議長声明を認識しつつ、オーストラリア、カナダ、欧州連合、フランス、ドイツ、イタリア、日本、韓国、シンガポール、英国及び米国を代表する世界の首脳は、開放性、透明性及び相互主義を基礎とし、A I 安全性の科学を前進させるための国際的な連携及び協力の重要性を確認する。我々は、責任あるA I イノベーションの促進において安全性が重要な要素であることを確認する。
2. 我々は、商業的及び公的に利用可能なA I システムのためのA I 安全性を前進させるために、A I 安全性の研究、評価及び／又はガイダンス策定を促進する、A I セーフティ・インスティテュートを含む、公的及び／又は政府が支援する機関を創設又は拡大するための共同作業を賞賛する。
 - 2.1 我々は、A I 安全性に関する政策取組に情報を提供するための、信頼でき、学際的で再現性のある一連のエビデンスの必要性を認識する。我々は、最終的にA I の開発及び導入の便益が世界中で公平に共有されるよう、科学的探求の役割とそのような探求を前進するための国際的連携の利点を認識する。
 - 2.2 我々は、国際A I 安全性報告書などの評価を通じて共通の科学的理解を活用及び促進し、適切な場合には我々それぞれの政策を誘導及び整合させ、また、我々のガバナンス枠組に沿って、安全、安心で信頼できるA I イノベーションを可能にするとの意図を確認する。
 - 2.3 我々は、我々の技術的方法論と全般的なアプローチにおける補完性と相互運用性を促進するよう努めることを含め、A I の安全性の側面に関する、共通の国際的、科学的理解を醸成するための手段を取るという共通の意図を表明する。

- 2.4 これらの手段は既存の取組を活用すること、研究、評価及びガイダンス能力を相互に強化すること、適切な場合には、能力、限界及びリスクに関するものを含むモデルに関する情報を共有すること、A Iによる害や安全性インシデントをモニタリングすること、適切な場合には、評価、データセット及び関連する基準を交換又は共同で作成すること、A I安全性の科学を推進するための共有の技術リソースの構築、この分野における適切な研究安全性の慣例を推進することを含む。
 - 2.5 我々は、効率を最大化し、優先順位を明確にし、進捗状況を報告し、我々の成果の科学的厳密性と堅牢性を高め、国際基準の策定及び採択を促進し、A I安全性に関するエビデンスに基づくアプローチの前進を加速させるために、我々の努力を連携する。
3. 我々は、A I安全性の科学の前進を加速するため、主要なパートナー間の国際的ネットワークを発展させるという共通の願いを表明する。我々は、これら及び関連する取組について、緊密な将来の協力、対話及びパートナーシップを期待する。