



広島AIプロセスの検討状況と今後の進め方

令和5年11月
国際戦略局

- 10月30日、AIに関するG7首脳声明を発出
 - AI開発者向けの国際指針及び国際行動規範について首脳が歓迎し正式公表
 - G7外を含めた地域へのアウトリーチとマルチステークホルダーとの協議を指示
- 首脳指示を踏まえ、国内マルチステークホルダーに対する協議(パブコメ)とG7外へのアウトリーチを実施
- これらの結果も踏まえ、全てのAI関係者向けの国際指針等の策定作業を引き続き実施し、年末までに再度閣僚級会合(オンライン)を開催し、包括的政策枠組として成果をとりまとめ、G7首脳に報告予定

G7プロセス

9月7日
閣僚級会合
(オンライン)

- ・閣僚声明とりまとめ

10月9日
非公式会合
@IGF京都
2023

- ・指針パブコメ案に合意

10月30日
G7首脳声明

- ・AI開発者向け国際指針及び国際行動規範の公表
- ・アウトリーチ及び協議の実施の指示 等

12月
閣僚級会合
(オンライン)

- ・広島AIプロセス包括的政策枠組をとりまとめ、首脳に報告

拡大プロセス

10月9日
マルチステークホルダー
ハイレベル会合
@IGF京都2023

- ・広島AIプロセスの議論を共有
- ・マルチステークホルダー協議の実施

G7外への
アウトリーチの実施

- ・英国安全性サミット(11月1・2日)
- ・GPAIサミット(12月13日)等、国際会議の活用も検討

マルチステークホルダー
協議の実施

パブコメ実施

- AI開発組織向けの国際指針について、10月9日のIGFでのG7非公式会合で基本合意。(EU、米、日本がそれぞれパブコメ)
- 10月30日、広島AIプロセスに関するG7首脳声明を発出し、AI開発組織向けの国際指針と国際行動規範について歓迎し、公表。

1. 高度なAIシステムの市場投入前及び、高度なAIシステムの開発を通じて、AIライフサイクルにわたるリスクを特定、評価、低減するための適切な対策を実施する

(行動規範例) 市場投入前の「レッドチーミング」などの内部および独立外部テストによるリスクの特定と低減

リスク例：化学・生物兵器の開発等に係るハードルを下げるリスク、有害な偏見や差別を社会等にもたらすリスク、偽情報助長やプライバシー侵害など民主主義的価値や人権に対するリスク

2. 市場投入後に脆弱性、インシデント、悪用パターンを特定し、低減する

(行動規範例) コンテストや賞金などを活用した、第三者および利用者による問題や脆弱性の発見と報告の促進

3. 十分な透明性の確保や説明責任の向上のため、高度なAIシステムの能力、限界、適切・不適切な利用領域を公表する

(行動規範例) 安全性・セキュリティ・社会や人権に対するリスクに関する評価、AIモデルの能力や限界等を含んだ透明性報告書や使用説明書の公表

4. 産業界、政府、市民社会、学术界を含む関係組織間で、責任ある情報共有とインシデント報告に努める

(行動規範例) 安全性・セキュリティ・信頼性を確保するため、情報共有のための基準・メカニズム・ベストプラクティスを開発し採用

5. リスクベースのアプローチに基づいたAIのガバナンスとリスク管理ポリシーを開発、実践、開示する。特に高度AIシステムの開発者向けの、プライバシーポリシーやリスクの低減手法を含む。

(行動規範例) 個人データ、ユーザーのプロンプトや出力を含めたプライバシーポリシーの開示
職員が自らの責務や組織のリスク管理慣行を熟知するための方針・手順・訓練の確立

6. AIのライフサイクル全体にわたり、物理的セキュリティ、サイバーセキュリティ及び内部脅威対策を含む強固なセキュリティ管理措置に投資し、実施する

(行動規範例) 情報セキュリティに関する安全運用措置等による「モデルウェイト」やアルゴリズムの保護
最も貴重な知的財産や企業秘密を保護するための強固な内部脅威検知プログラムの確立

7. AIが生成したコンテンツを利用者が識別できるように、電子透かしやその他の技術等、信頼性の高いコンテンツ認証および証明メカニズムを開発する。またその導入が奨励される。

(行動規範例) 電子透かしや証明システムなど、AI生成コンテンツであることを利用者が判断できるためのツールやAPIの開発
AIと接していることを利用者が認知できるようなラベリング表示メカニズムの導入

8. 社会、安全、セキュリティ上のリスクの低減のための研究を優先し、効果的な低減手法に優先的に投資する

(行動規範例) 民主的価値の確保や人権の尊重等に関する研究の実施、協力や投資
環境及び気候への影響を含むリスク低減ツールや積極的リスク管理作業への投資

9. 気候危機、健康・教育などの、世界最大の課題に対処するため、高度なAIシステムの開発を優先する

(行動規範例) 国連SDGsの進捗を支援するためのAI開発を支援

10. 国際的な技術標準の開発と採用を推進する

(行動規範例) 電子透かしを含む国際的な技術標準とベストプラクティスの開発や利用に貢献

11. 適切なデータ入力措置と個人情報及び知的財産の保護を実施する

(行動規範例) プライバシーや知的財産を尊重するための安全措置の実施
適用される法的枠組みの順守

- 12月を目途にG7デジタル技術閣僚級会合をオンラインで開催。
- パブコメ結果等を踏まえ、閣僚級会合で、以下4点を内容とする「広島AIプロセス包括的政策枠組み」に合意。
 1. 生成AIに関する優先的な課題、リスク及び機会に関する分析
 2. AI開発者を含む全てのAI関係者向けの国際指針
 3. AI開発者向けの行動規範
 4. 偽情報対策に資する研究の促進等のプロジェクトベースの協力
- 広島AIプロセスを更に前進させるための作業計画を年内に策定。
- 上記成果を首脳に報告。