

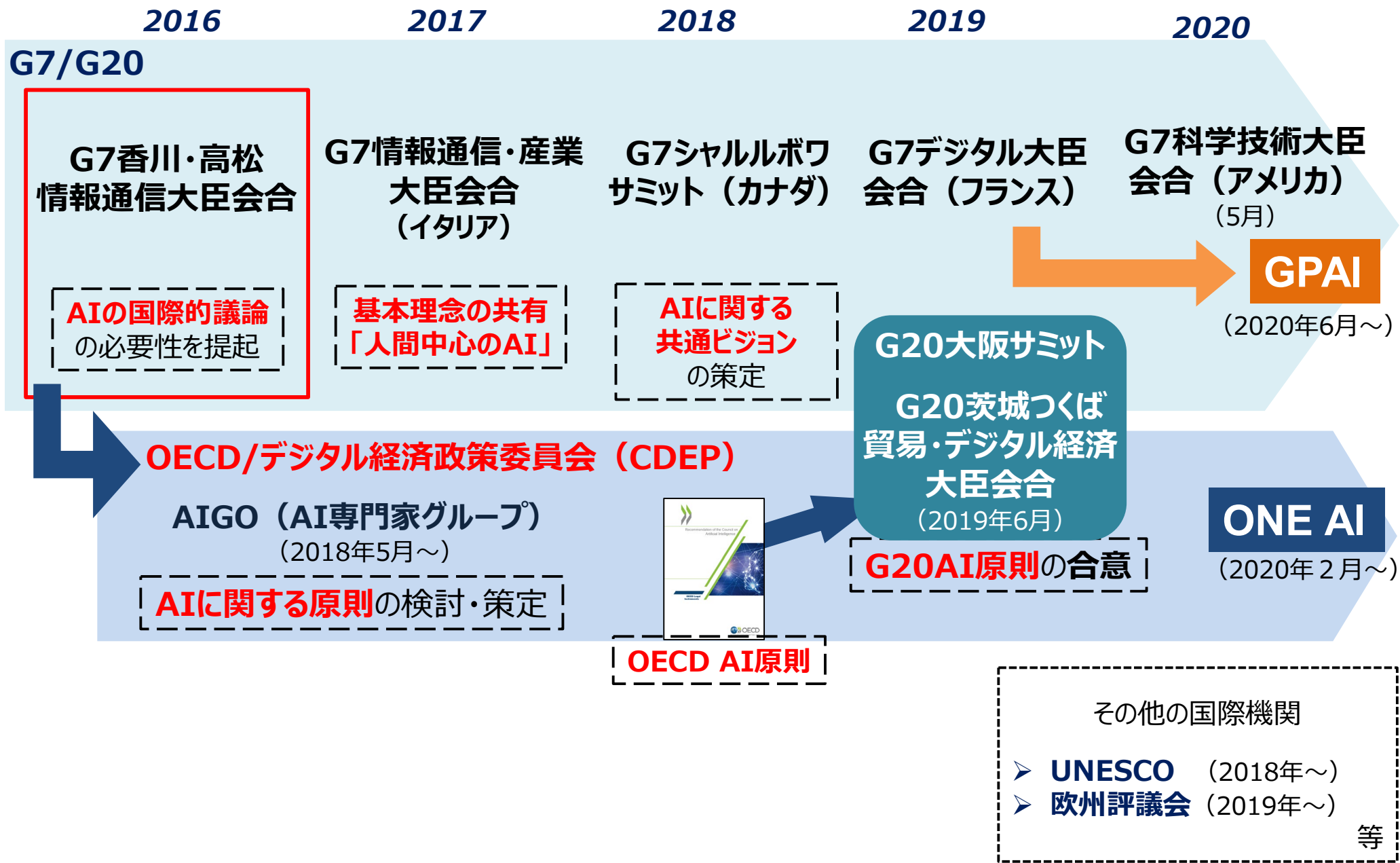


広島AIプロセスの成果と意義

(一財) 日本ITU協会理事長 吉田博史

A I に関する議論の国内外における経緯の概観

1



【経緯】

2022年11月30 日
Open AI社 ChatGPT 3.5 プロトタイプ リリース

2023年3月15日
Chat GPT 4 リリース

2023年4月10日
サム・アルトマンCEO 岸田総理を訪問

2023年4月29～30日
G7デジタル・技術大臣会合 （群馬・高崎）

2023年5月19～21日
G7広島サミット



【広島サミット首脳コミュニケ（抜粋）】

生成AIの機会と課題を早急に把握する必要性を認識し、OECDやGPAIにおける取組を奨励。関係閣僚に対して、生成AIに関する議論のために、包摂的な方法で、OECDやGPAIと協力しつつ、**G7作業部会を通じた、広島AIプロセスを年内に創設するよう指示**。この議論には、ガバナンス、知的財産権保護、透明性促進、偽情報への対策及びこれらの技術の責任ある活用といったテーマを含み得る。

- 2023年5月のG7広島サミットの結果を受けて、**生成AIに関する国際的なルールの検討を行うため、「広島AIプロセス」**を立ち上げ。
- 10月30日に「**広島AIプロセスに関するG7首脳声明**」が発出され、**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指針及び国際行動規範**を公表。
- 12月1日にG7日本議長国下での取組の成果として、「**広島AIプロセス包括的政策枠組**」及び「**広島AIプロセスを前進させるための作業計画**」をとりまとめ。12月6日に発出されたG7首脳声明で**これらの成果をG7首脳が承認**。
- 2024年3月にG7イタリア議長国下で採択された「G7産業・技術・デジタル閣僚宣言」においても**広島AIプロセスの成果の前進にG7が継続的にコミット**することを確認。
- 5月のOECD閣僚理事会において、49の国・地域の参加を得て、広島AIプロセスの精神に賛同する国々の自発的な枠組みである「**広島AIプロセス フレンズグループ**」を立ち上げ。

日本議長国（2023年）

イタリア議長国（2024年）

G7デジタル
閣僚級会合

閣僚声明とりまとめ

G7首脳声明高度なAIシステムを開発
する組織向けの広島プロ
セス国際指針及び国際
行動規範の公表G7首脳声明12月の閣僚級会合
の成果を承認

国際行動規範の実践

G7産業・技術・
デジタル大臣会合

5月

9月

10月

12月

3月

5月

G7広島サミット広島AIプロセス
立ち上げを指示マルチステークホルダー
ハイレベル会合（IGF）

広島AIプロセスの議論を共有

G7デジタル
閣僚級会合国際指針、国際行動規
範、プロジェクトベース協
力等を含む
包括的政策枠組及び今
後の作業計画をとりまとめOECD閣僚理事会広島AIプロセスフレンズグループの
立ち上げ

賛同国の拡大

- 令和5年12月1日（金）、総務省、経産省及びデジタル庁共同で「G7デジタル・技術大臣会合」を開催。G7構成国・地域のほか、関係国際機関が参加。
- 広島AIプロセス（議長：鈴木総務大臣）及びDFFT（議長：河野デジタル大臣）について議論を行い、成果文書として、「広島AIプロセス G7デジタル・技術閣僚声明」及び「DFFT の具体化に関する閣僚声明」が採択。

広島AIプロセス G7デジタル・技術閣僚声明における主な成果① - 広島AIプロセス包括的政策枠組み

- ◆ 生成AI等の高度なAIシステムへの対処を目的とした初の国際的枠組みとして「**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G7の共通理解に向けたOECDレポート

- 優先的な課題・リスクとして、透明性、偽情報、知的財産権、プライバシーと個人情報保護、公正性、セキュリティと安全性等を例示。機会として、生産性・イノベーション促進、ヘルスケア改善、気候危機解決への貢献等。

2. 全てのAI関係者向け**広島プロセス国際指針**

- 「AI開発者向けの国際指針」（2023年10月30日公表）の11項目が高度なAIシステムの設計、開発、導入、提供及び利用に関わる全ての関係者に適宜適用し得ることを確認。
 - ・ 偽情報の拡散等のAI固有のリスクに関するデジタルリテラシーの向上、脆弱性の検知への協力と情報共有等、利用者に関わる内容を12番目の項目として追加。

3. 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**

- 10月30日に公表した国際行動規範を支持する声明を発出している組織をG7として歓迎。
- 幅広い支持を得るために、より多くの組織への働きかけを継続することを確認。

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**

- OECD,GPAI及びUNESCO等が実施する「生成AI時代の信頼に関するグローバルチャレンジ」の取組を歓迎。
- GPAI東京センターを含め生成AIに関するGPAIプロジェクトを歓迎。（コンテンツ認証・来歴管理メカニズム等）

広島AIプロセス G7デジタル・技術閣僚声明における主な成果② - 広島AIプロセス推進作業計画

◆ G7として、以下の項目の「**広島AIプロセスを前進させるための作業計画**」についても合意。

1. 広島プロセス国際指針及び行動規範への**賛同国増加に向けたアウトリーチ**
2. 企業等による**国際行動規範への支持拡大**及び**企業等による国際行動規範履行確保のためのモニタリングツール**の導入に向けた取組の実施
3. グローバル・チャレンジやその他の潜在的な機会を通じた、**OECD、GPAI、UNESCOとのプロジェクトベースの協力の継続**

これらの他、以下の取組を推進。

- 関連国の政策動向及び国際行動規範にコミットする組織のリストに関する最新情報等を提供する**広島AIプロセス専用ウェブサイト**の立ち上げ
- マルチステークホルダーコミュニティとの対話促進を通じた、広島AIプロセスの成果の推進
- **OECDに対して既存のAIの取組みにおいて広島AIプロセスの成果を考慮するよう奨励**
- **OECD、GPAI及び国連等の多国間の場**における協調と協力の強化による**広島AIプロセスの更なる前進**

- 安全、安心、信頼できるAIの実現に向けて、AIライフサイクル全体の関係者それぞれが異なる責任を持つという認識の下、12の項目を整理。
- 「AI開発者向けの広島プロセス国際指針」の11の項目が、高度なAIシステムの設計、開発、導入、提供及び利用に関わる全ての関係者に適宜適用し得るものとして整理した上で、偽情報の拡散等のAI固有リスクに関するデジタルリテラシーの向上や脆弱性の検知への協力と情報共有等、利用者に関わる内容が12番目の項目として追加。

全てのAI関係者向けの広島プロセス国際指針の12項目

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。
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。
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。
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。
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。
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。
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。
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。
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。
10. **国際的な技術標準の開発と採用を推進**する
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。
12. **偽情報の拡散等のAI固有リスクに関するデジタルリテラシーの向上や脆弱性の検知への協力と情報共有等**、高度なAIシステムの**信頼でき責任ある利用を促進し、貢献**する。

- 10月30日、広島AIプロセスに関するG7首脳声明を発出し、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について歓迎し、公表。

1. 高度なAIシステムの**開発・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、評価し、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. 開発・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、学術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