ブリュッセル効果への対応 [第6回]
企業にとどまらぬAI透明性義務の射程

2026年1月6日(火) 14:00-15:00

【プログラム】

14:00-14:05	開会挨拶：福田和樹（内閣府）
14:05-14:15	EU-AI法50条の概説：工藤郁子（大阪大学）
14:15-14:30	EU AI Officeが進めるCoP策定プロセスについて: 実積寿也（中央大学）
14:30-14:55	パネルディスカッションとQ&A パネリスト： 工藤郁子（大阪大学） 実積寿也（中央大学） 村上明子（日本AIセーフティ・インスティテュート） 佐渡島庸平（株式会社コルク） 三宅陽一郎（東京大学） 司会：江間有沙（東京大学東京カレッジ）
14:55-15:00	閉会挨拶：村上明子（日本AIセーフティ・インスティテュート）

UTokyo | TOKYO COLLEGE

東京大学未来ビジョン研究センター
Institute for Future Initiatives
The University of Tokyo

AI Center UTokyo

大阪大学

ブリュッセル効果への対応6：企業にとどまらぬAI透明性義務の射程

EU AI法 50条の概説

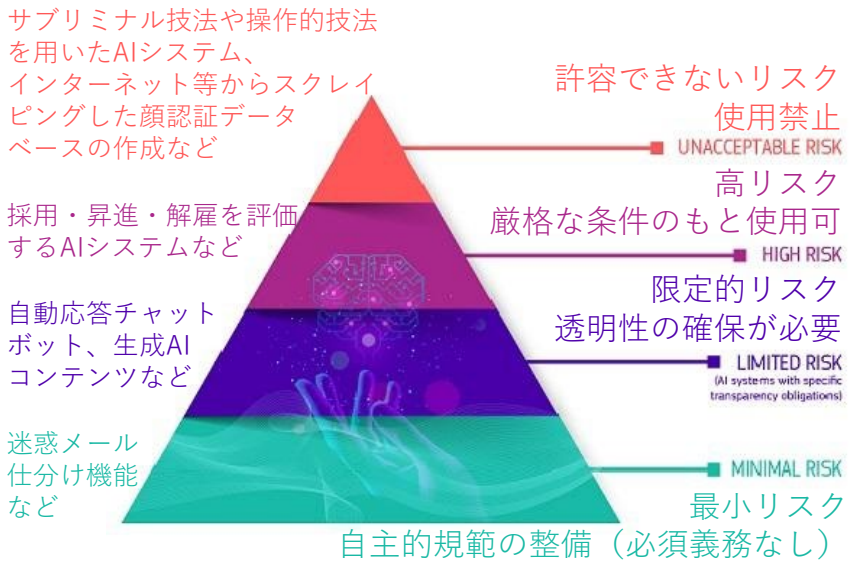
大阪大学 社会技術共創研究センター 特任准教授
工藤郁子
2026年1月6日

大阪大学 社会技術共創研究センター
Research Center on Ethical, Legal and Social Issues

ELSI

EU AI法

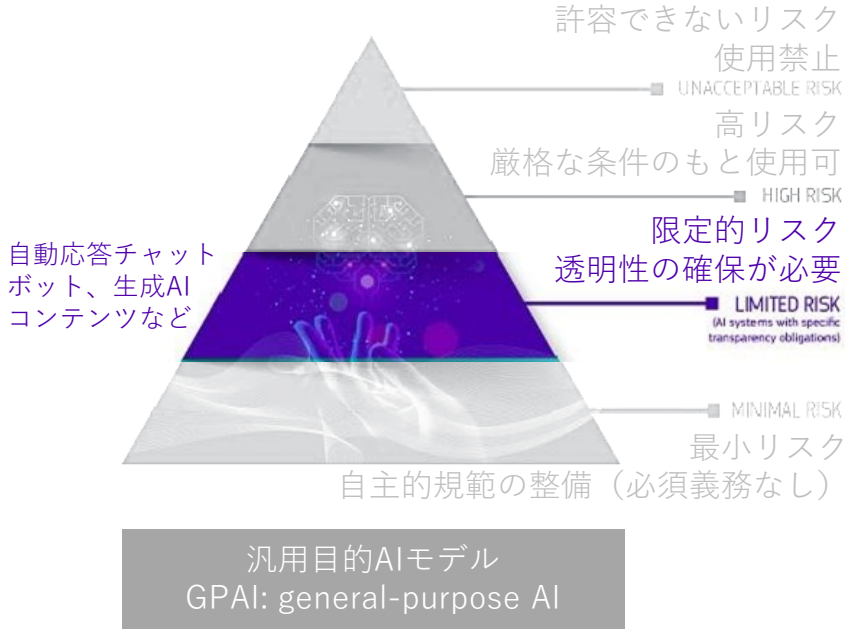
- ハードロー：世界で初めて包括的にAIを規制する法律
- リスクベースアプローチ：リスクの大きさに合わせた対応(右図参照)
  - 「リスク」とは、害が発生する確率とその害の重大性の組み合わせ(3条2号)
- 域外適用：日本企業も対象(2条)
  - EU域内でAIシステムをEU市場に流通させる/サービスを提供するプロバイダー(EUに所在しているかは問わない)
  - アウトプットがEU域内で利用される場合、第三国に所在するAIシステムのプロバイダーとデプロイヤー
- 違反には多額の罰金(99～101条)
- 本ウェビナーシリーズ前回まで(ブリュッセル効果への対応1～5)は、汎用目的AI（GPAI）モデルの規律を対象(51～56条)



EUのウェブサイトの資料をもとに発表者作成
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

EU AI法：50条の限定的リスク

- 今回の対象は「限定的リスク」の透明性の確保
  - 典型例：AIが合成・操作した動画だと表示する等
  - なお、「高リスク」なAIシステムが、50条にも該当する場合、透明性義務も重ねて負う(50条6項)
- さまざまな例外も認められている
  - 実在の人間・物・場所・出来事と類似し、本物・真実だと誤認されうる「ディープフェイク」について(3条60号)、明らかに風刺やフィクションである場合、作品の鑑賞・享受を妨げない適切な開示方法で足りる(50条4項)等
- 具体的な部分は、Code of Practice: CoP(行動規範)を策定することになっている(50条7項)
  - CoPは、ガイドラインで法的拘束力はない
  - しかし、デファクトスタンダードとして機能することも想定される (CoPとは異なる独自運用を選択する場合、EU AI Officeとの個別調整が必要となる可能性)



EUのウェブサイトの資料をもとに発表者作成
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

EU AI法：プロバイダーとデプロイヤー

- 限定的リスクに関するAI透明性義務を負うのは、プロバイダーとデプロイヤー
  - プロバイダー：AIシステムの開発者や、自らの名称や商標で保有するAIシステムをEU市場に流通させる/サービスを提供する者（3条3号）
  - デプロイヤー：自らの権限に基づき業務でAIシステムを利用する者（3条4号）
- 画像・音声・動画を生成するAIシステムの開発者・提供者だけでなく、例えば、AIシステムを業務利用した、デザイナー、マーケター、YouTuber、ゲームクリエイター、記者なども50条のデプロイヤーとして対象になりうる
  - 中小企業や個人事業主なども対象になる点に注意

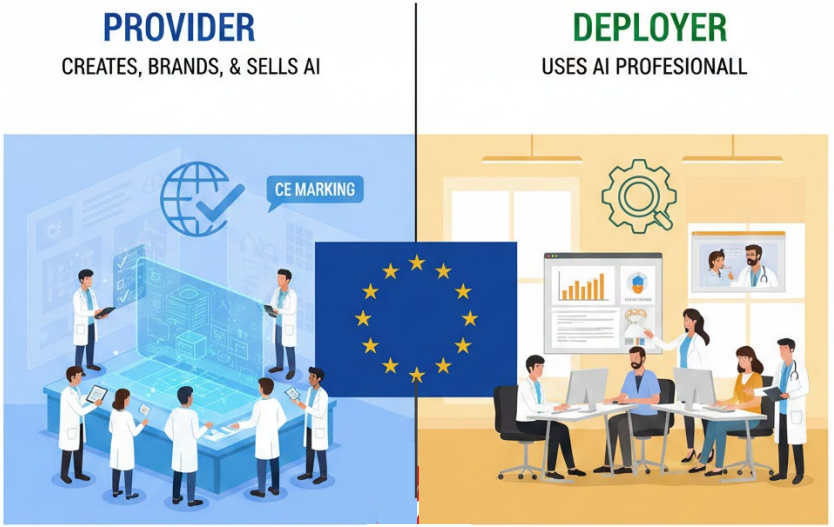


Image created with Gemini

EU AI法：50条のAI透明性義務

条文	対象	義務
50条1項	• 典型例：**対話型AI** • 人間と直接対話することを目的としたAIシステム	• **プロバイダー**が、利用者に対して、AIシステムと対話していることが通知されるように、AIシステムを設計・開発する義務
50条2項	• 典型例：**コンテンツ生成AI** • 合成音声、画像、動画、テキストコンテンツを生成するAIシステム（汎用目的AIを含む）	• **プロバイダー**が、生成された出力について、機械可読形式で、人工的に生成・操作されたものであると検知可能にする義務 • **プロバイダー**が、コンテンツの特殊性、実装コスト、最新技術を考慮し、技術的に実行可能な限り、その技術的措置について、効果的、相互運用可能、堅牢、信頼できるよう確保する義務
50条3項	• 典型例：**感情認識AI、顔認証AIカメラ** • 感情認識システム、生体認証分類システム	• **デプロイヤー**が、対象となる自然人に、AIシステムが動作していることについて通知等する義務
50条4項	• 典型例：**ディープフェイク** • 画像、音声、動画コンテンツを生成・操作してディープフェイクを作成するAIシステム	• **デプロイヤー**が、当該コンテンツは人工的に生成・操作されたものであると開示する義務 • 例外：明らかに風刺やフィクションである場合
	• 典型例：**選挙情報、災害情報** • 公衆への情報提供を目的とし、公共の利益に関するテキストを生成・操作するAIシステム	• **デプロイヤー**が、当該テキストが人工的に生成・操作されたものであることを開示する義務 • 例外：人間によるレビューや編集管理プロセスを経ており、自然人・法人がコンテンツの公表に編集責任を有する場合等
50条5項	• 50条1～4項は、遅くとも最初の対話または露出の時点までに、利用者に対して、明確で認識可能な方法で提供されなければならない	

CoP：2つのワーキンググループ(WG)

条文	対象	義務
50条1項	• 典型例：対話型AI • 人間と直接対話することを目的としたAIシステム	• プロ こと
50条2項	• 典型例：**コンテンツ生成AI** • 合成音声、画像、動画、テキストコンテンツを生成するAIシステム（汎用目的AIを含む）	• プロ プロ を考慮し、技術的に実行可能な限り、その技術的措置について、効果
50条3項	• 典型例：感情認識AI、顔認証AIカメラ • 感情認識システム、生体認証分類システム	• デフ いる
50条4項	• 典型例：**ディープフェイク** • 画像、音声、動画コンテンツを生成・操作してディープフェイクを作成するAIシステム	• プ ものであることを開示する義務 • 例外
	• 典型例：**選挙情報、災害情報** • 公衆への情報提供を目的とし、公共の利益に関するテキストを生成・操作するAIシステム	• デフ の • 人 然人
50条5項	• 50条1～4項は、遅くとも最初の対話または露出の時点までに、利用者に対して、明確で認識可能な方法で提供されなければならない	

CoP WG1
Marking and detection techniques for providers (Art. 50(2))

CoP WG2
Disclosure of ‘deep fakes’ and certain AI-generated text (Art. 50(4))

CoP WG1+WG2
Interplay issues and horizontal requirements (Art. 50(5))

CoP：論点例

条文	対象	論点例
50条1項	• 典型例：対話型AI • 人間と直接対話することを目的としたAIシステム	• **プロバイダー**が、利用者に対して、AIシステムと対話していることが通知されるように、AIシステムを設計・開発する義務
50条2項	• 合成音声、画像、動画、テキストコンテンツを生成するAIシステム（汎用目的AIを含む） • **プロバイダー**が、生成された出力について、機械可読形式で、人工的に生成・操作されたものであると検知可能にする義務など	• **各マーキング技術（メタデータ識別子、暗号署名、電子透かし等）は単独では不十分だが、どんな組み合わせがいいか？** • **中小企業やスタートアップにとって、実装が容易で費用対効果の高い方法は？**
50条3項	• 典型例：感情認識AI、顔認証AIカメラ • 感情認識システム、生体認証分類システム	• **デプロイヤー**が、対象となる自然人に、AIシステムが動作していることについて通知等する義務
50条4項	• 画像、音声、動画コンテンツを生成・操作してディープフェイクを作成するAIシステム • **デプロイヤー**が、当該コンテンツは人工的に生成・操作されたものであると開示する義務など	• **作品の鑑賞を妨げずに、コンテンツの透明性を確保するには？** • **「明らかに」フィクションである場合とは、どんなもの？** • **過小でも（メタデータのみ）過多でも（AI生成ラベルのスパム）、ユーザーは無視するだろうが、どうすればいいか？**
	• 公衆への情報提供を目的とし、公共の利益に関するテキストを生成・操作するAIシステム • **デプロイヤー**が、当該テキストが人工的に生成・操作されたものであることを開示する義務	• **開示義務の免除について、人間による効果的なレビュープロセス・編集管理・編集責任をどのように確保するか？** • **小規模デプロイヤーと大規模デプロイヤーで課題が異なるか？**
50条5項	• 最初の露出の時点までに、利用者に対して提供	• **情報を適時に提供するためのグッドプラクティスと措置は？**

Guidelines and Code of Practice on Transparent AI Systems

中央大学　実積寿也

Prepared by T.Jitsuzumi for 1/6/2026.

AI法50条

1. 位置づけ・意味
   - 第50条は 第IV章「特定AIシステムの透明性義務」 の中核条文で、高リスクかどうかに関わらず、一定のAIシステムに横断的な透明性義務を課す規定。
   - 人々が「AIと対話している／AI生成・操作コンテンツを見ている」ことを認識できるようにし、情報の真正性・信頼性を確保することが目的。

2. 対象となるAIシステム
   - 人と直接やりとりするAIシステム（チャットボット等）
   - 合成音声・画像・動画・テキストを生成するAI（GPAIを含む）
   - 感情認識システム・バイオメトリック分類システム
   - ディープフェイク（人物・出来事等を本物らしく見せる画像・音声・動画）
   - 公共の関心事項について一般に公開されるAI生成・操作テキスト（ニュース等）

3. プロバイダー／デプロイヤーの主な義務
   - 対話型AI：利用者が「AIとやりとりしている」ことを通知（“誰が見ても明らか”なら例外）。
   - 合成コンテンツ生成AI：出力を機械可読な形式でマーキングし、検知可能にする（有効性・相互運用性・堅牢性・信頼性、コスト、state of the artを考慮）
   - 感情認識・バイオメトリック分類：自然人にシステム使用を通知し、GDPR等に対応。
   - ディープフェイク：画像・音声・動画が人工的に生成・操作されたことを明示（芸術・フィクション等では「作品の享受を妨げない形」での開示に限定）。
   - 公共的テキスト：ニュース等として公開されるAI生成テキストは、その旨を開示（ただし人間による実質的レビューと編集責任があれば免除）。

4. 情報提供の方法とCoPとの関係
   - 情報は「明確かつ識別可能な形」で、遅くとも最初の接触・曝露時までに 提供し、障がい者へのアクセシビリティ要件にも適合させる。
   - 第50条自体は法的に拘束力のある義務条文であり、その実務的実装を支援するために、AIオフィスが 検知・ラベリングに関するCoPを策定し、欧州委員会が承認できる枠組みを規定。

Prepared by T.Jitsuzumi for 1/6/2026.

プロバイダーとデプロイヤーの定義、オープンソースの取扱い

- AI法第3条
  - (3) ‘provider’ means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge;
    - 「プロバイダー」とは、AIシステムまたは汎用AIモデルを開発する、またはAIシステムまたは汎用AIモデルを開発させ、自己の名または商標の下で、有償・無償を問わず、市場に流通させる、またはAIシステムをサービスとして提供する自然人、法人、公的機関、機関その他の団体をいう。
  - (4) ‘deployer’ means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity;
    - 「デプロイヤー」とは、個人または法人、公的機関、政府機関その他の団体であって、その権限下でAIシステムを使用する者をいう。ただし、当該AIシステムが個人の非職業的活動において使用される場合は除く。
  - 企業規模による例外措置はなく、SMEであっても対象
  - 個人の場合も、職業としてAIシステムを開発・利用する場合には対象

- AI法2条　適用範囲
  - 12. This Regulation does not apply to AI systems released under free and open-source licences, unless they are placed on the market or put into service as high-risk AI systems or as an AI system that falls under Article 5 or 50.
    - 本規則は、フリーかつオープンソースのライセンスに基づき公開されたAIシステムには適用されない。ただし、当該システムが高リスクAIシステムとして、または第5条もしくは第50条に該当するAIシステムとして市場に流通させる場合、もしくは使用に供する場合を除く。

Prepared by T.Jitsuzumi for 1/6/2026.

罰則（Chapter XII）

- 1）事業者（provider / deployer）に対する制裁（加盟国法で執行）
  - 第50条（providers/deployersの透明性義務）違反は、最大で「1,500万ユーロ」または「全世界年間売上高の3%」のいずれか高い方の行政罰の対象（Article 99(4)(g)）。
  - SME（中小企業）上限の特則：SME（スタートアップ含む）の場合、上記の罰金上限は「金額上限」か「売上高比率上限」のうち低い方になります。
  - 量定要素：個別事案では、違反の性質・重大性・継続期間・影響、事業者の規模や売上、当局への協力等を考慮する枠組みが置かれています。

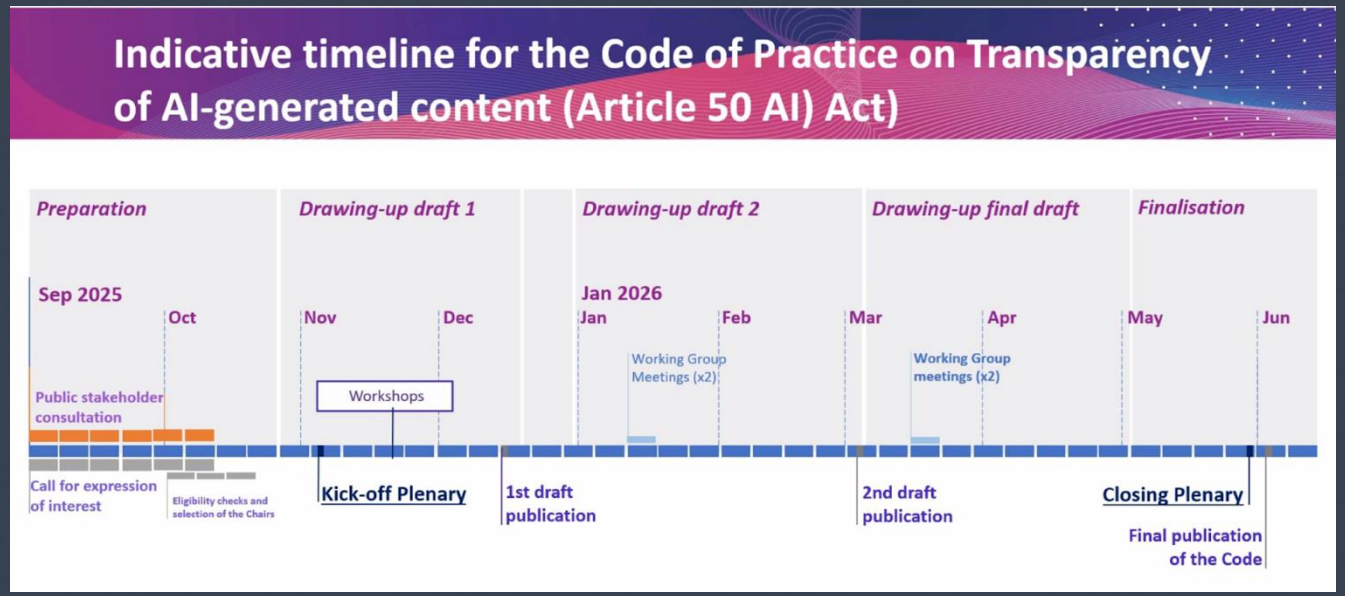
- 2）EU機関等（Union institutions）には別枠の上限
  - EU機関・機関等に対しては、EDPS（欧州データ保護監察官）が行政罰を科し得る別条（Article 100）があり、本規則上の要件・義務違反（＝第50条違反も含み得る）は、最大75万ユーロ。

Prepared by T.Jitsuzumi for 1/6/2026.

AI生成コンテンツ透明性CoP検討の全体枠組み

1. 背景・目的
   - 根拠：EU AI Act 第50条（AI生成・操作コンテンツの透明性義務）
   - 目的：
     - 生成AIプロバイダーによる機械可読なマーキング＆検知（Art.50(2)）
     - デプロイヤーによるディープフェイク・公共的テキストの開示（Art.50(4)）を実務的に支援する任意のCode of Practice（CoP）を策定

2. 構造：2つのWG＋横断要件
   - WG1（プロバイダーWG）→ 生成AIシステムのマーキング＆検知技術・標準・ガバナンス（Art.50(2)）
   - WG2（デプロイヤーWG）→ メディア・プラットフォーム等によるラベリング実務・ユーザーへの開示（Art.50(4)）
   - 合同・横断WG（Art.50(5)）→ 表示形式、タイミング、アクセシビリティなど共通の開示要件を整理



AI ASSISTED
AI-GENERATED CONTENT
EU AI Act · Art. 50

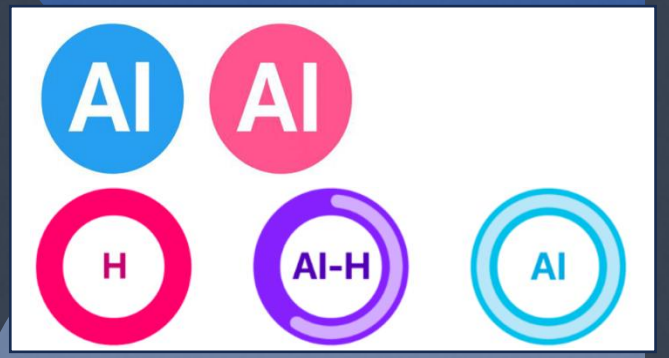
Prepared by T.Jitsuzumi for 1/6/2026.

First draft（Dec. 17, 2025）

1. プロバイダー向け：技術的なマーキングと検出（Section 1）
   - AIシステム提供者は、出力がAI生成であることを機械可読な形式で識別可能にする必要がある。
   - 多層的アプローチの採用：単一の技術では不十分であるとし、以下の技術を組み合わせた多層的な実装を要求。
     - メタデータ埋め込み：来歴情報やAIシステムの署名をメタデータに格納。
     - 電子透かし：除去が困難な非可視のウォーターマーク。
     - フィンガープリント/ロギング：ハッシュ値などを用いた照合手段。
   - 検出ツールの提供：ユーザーや第三者がコンテンツの真偽を検証できる検出インターフェースやAPIを無償で提供。
   - 品質要件：有効性、堅牢性（攻撃への耐性）、相互運用性を満たす必要あり。

2. デプロイヤー向け：ディープフェイクとテキストの開示・表示（Section 2）
   - AIシステムを利用してコンテンツを公開する者は、それがAIによって生成・操作されたものであることを自然人に開示する必要がある。
   - 共通の分類（タクソノミー）　：以下の2つを区別して管理・開示。
     - 完全なAI生成（Fully AI-generated）：プロンプト等からAIが自律的に生成。
     - AI支援（AI-assisted）：AIによる操作が意味や真実性に実質的に影響。
   - 共通アイコンの使用：欧州共通のアイコンが完成するまでの間、「AI」という文字を含む暫定的なアイコンを視認可能な位置に表示。
   - モダリティ別の対応：動画（ディープフェイク）、音声、テキスト
   - 例外措置：　芸術・風刺・創作作品については、作品の鑑賞を妨げない範囲で、控えめな開示（non-intrusive position）を容認。

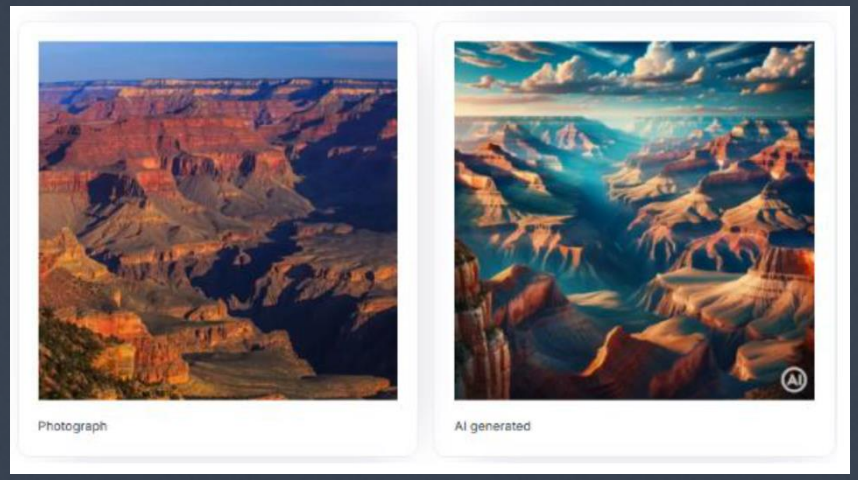
- 2026年1月23日までフィードバックを受け付け、さらなる改訂が行われる予定



Prepared by T.Jitsuzumi for 1/6/2026.

モダリティ別の対応

- 画像



- 動画（ディープフェイク）
  - 開始時の免責事項（ディスクレーマー）や、動画内へのアイコン常時表示など
- 音声
  - 音声によるディスクレーマー（冒頭、中間、終了時など）
- テキスト
  - 公益に関する情報を通知する目的で公開されるテキストについては、犯罪捜査目的等の法執行要請に基づくもの、もしくは人間による編集責任があるものでない限り、AI生成であることを開示する必要がある

Prepared by T.Jitsuzumi for 1/6/2026.

Section 1（AIシステムプロバイダー向けルール）

Commitment 1: Multi-layered Marking of AI-Generated Content

- Measure 1.1：複数層の機械可読マークで自社AIによる出力結果をマーキング。マーキングはバリューチェーンの複数の段階で付与されたり、第三者によって付与される場合もある。
  - Sub-measure 1.1.1：メタデータ埋込み可能なコンテンツには来歴、システム署名、処理種別を電子署名として付加
  - Sub-measure 1.1.2：加工・攻撃耐性を持つ不可視な電子透かしを付与
  - Sub-measure 1.1.3：（必要あれば）フィンガープリント／ログで「自社AIの出力か否か」を確認可に
- Measure 1.2：モダリティ別の追加マーキング手段を整備
  - Sub-measure 1.2.1：メタデータ埋込みが不可能な場合は電子署名付き来歴証明を付与
  - Sub-measure 1.2.2：マルチモーダルの場合、各モダリティのマーキングを相互同期
- Measure 1.3：オープンウェイトAIモデル提供者は構造的マーキング技術をウェイトに埋め込み、下流事業者によるMesure2.2遵守を支援
- Measure 1.4：モデル提供者はモデル上市前に自社モデル出力用の機械可読マーキング技術を実装
- Measure 1.5：既存マークは保持（変換後も）し、除去・改竄禁止を規約等で明記
- Measure 1.6：（技術的に可能な範囲で）コンテンツの来歴情報を記録・埋込み
- Measure 1.7：AIデプロイヤーのコンテンツ生成時点で“可視ラベル”を付与する機能を提供

Prepared by T.Jitsuzumi for 1/6/2026.

Section 1（AIシステムプロバイダー向けルール）

Commitment 2: Detection of the Marking of AI-Generated Content

- Measure 2.1：信頼性スコア付き検証API/UI（または一般利用可能な検出器）を無償提供
- Measure 2.2：モデル提供者は上市前に「自モデル出力の検出機構」を提供
- Measure 2.3：能動的マーキングに依存しない検出手段を実装。関連の技術開発の支援
- Measure 2.4：（技術的に可能な範囲で）検出・来歴結果に、人間が理解できる根拠説明を埋込み＋アクセシビリティ対応
- Measure 2.5：ツール選択・解釈のための文書/教材の提供、リテラシー向上の支援

Commitment 3: Measures to meet the Requirements for Marking and Detection Techniques

- Measure 3.1：有効性（低コスト、リアルタイム運用、コンテンツ品質維持、環境配慮）
- Measure 3.2：信頼性（偽陽性/偽陰性等で評価し、初見データで低誤判定を実証）
- Measure 3.3：堅牢性（加工・攻撃耐性）＋検出器・機能提供の際はセキュリティに配慮
- Measure 3.4：相互運用性（システム相互運用性確保、公開技術規格採用、標準化支援）
- Measure 3.5：研究投資を通じたAI利用検証技術の水準向上（電子透かし）

Prepared by T.Jitsuzumi for 1/6/2026.

Section 1（AIシステムプロバイダー向けルール）

Commitment 4: Testing, verification and compliance

- Measure 4.1：マーキング/検出のコンプライアンス枠組みを作成・実装・更新、市場監視当局からの要請があれば文書提出。中小企業（SME and SMC）には義務緩和あり
- Measure 4.2：上市前＋定期的に実環境で試験し、脅威モデリングで検出器を更新
- Measure 4.3：AIシステム作成やコンプライアンスの監督に携わる人員に訓練を実施。中小企業には義務緩和あり
- Measure 4.4：市場監視当局への協力、情報提供

Glossary

Term	Definition
Active marking	Addition or embedding of a marking to AI-generated or manipulated content such as a watermark or attached information such as a metadata entry. The purpose of this addition is to facilitate detection of this marking and provenance attribution of the AI-generated or manipulated content.
Active detection	Verification of markings such as watermarks or metadata markers that have been purposefully added by a provider of an AI system or model.
Adaptive threat modelling approach	A defensive measure in cybersecurity to continuously monitor and, if necessary, to adapt the security of a system.
Amortization attacks	A method in which an attacker performs one difficult or time-consuming task upfront and then re-uses that work to make many follow-up attacks much cheaper and faster.
API	Stands for Application Programming Interface, a machine-usable interface to an AI system or another software service from an AI system provider.
Digital signature	A cryptographic signature that enables verification of authenticity of the provider and integrity of the signed content.
Fingerprinting	Detection technique for image, video, audio, or text, based on either hashing or logging.
Forensic detection	Detection of AI-generated or manipulated content which does not depend on the presence of active AI marking. For example, a forensic method may attribute an image to an AI image generator using a signal characteristic in the image data or a machine learning model trained to distinguish AI-generated images from authentic ones.
Hashing / Perceptual Hashing	Reduction of audio or visual content to a short identifier for indexing. Can be used for a fast lookup for known content, i.e., a repository of hashes can be queried to find out whether content is known to have been AI-generated or manipulated.
Logging	Verbatim recording and indexing of content (usually text). Can be used for a fast lookup of known content, i.e., a repository of logged entries can be queried to find out whether content is known to have been AI-generated or manipulated.
Open-weight Model	A model where the underlying weights, code, and parameters are made publicly available.
Provenance Information	A digital record for a piece of content generated or manipulated by an AI system that shows its origin, how and when the content was generated or manipulated and processing steps applied to the content.
Shared verifier	A detector or verifier for markings originating from multiple providers of AI systems or models that generate or manipulate content.
Structural marking	An imperceptible watermark that is either embedded into a model during training or upon inference. This can be a technique to add a marking to an open-source model that can be downloaded from the internet. However, in this case its security is inherently limited because the watermarking key must at least implicitly be shipped with the model.
Synchronization of markings	Cross-referencing between markings in multimodal content. For example, a document consisting of a text and an image may contain a marking in the text that refers to the image, and a marking in an image that refers to the text, such that one cannot replace only the text or only the image without this affecting the integrity of the markings.
User	Either a deployer within the meaning of Article 3 (4) the AI Act or another person that is using the AI system of a provider or a person exposed to the content.
Watermark	A marker directly connected and interwoven within the content, typically through an imperceptible modification of the content, such that it is difficult to remove without affecting the fidelity of the content.

Prepared by T.Jitsuzumi for 1/6/2026.

Section 2（AIシステムデプロイヤー向けルール）

Commitment 1: Disclosure of Origin of AI-Generated and Manipulated Content based on a Common Taxonomy and an Icon

- Measure 1.1：法50条(4)対象をFully AI-generatedとAI-assistedに分類
- Measure 1.2：ディープフェイク等は共通アイコンを文脈に適した位置へ一貫表示
  - Sub-measure 1.2.1：EU共通アイコン確定まで、各国語2文字略語（AI/KI/IA等）の暫定アイコンの使用が可能
  - Sub-measure 1.2.2：“EU共通アイコン”のサポート
    - 自然人が認識可能、異言語対応、固定位置表示、コンテンツへの不干渉、音声コンテンツへの対応

Commitment 2: Compliance, training and monitoring

- Measure 2.1：アイコン適用実務を内部文書化し、具体例も含めて更新・運用
- Measure 2.2：企業規模・取扱コンテントに応じた担当者訓練の実施
- Measure 2.3：不正確ラベルの通報チャネル＋市場監視当局・プラットフォーム等との協力

Commitment 3: Ensure Accessible Disclosure for all Natural Persons

- Measure 3.1：ラベルのアクセシビリティ確保のための技術的・組織的支援

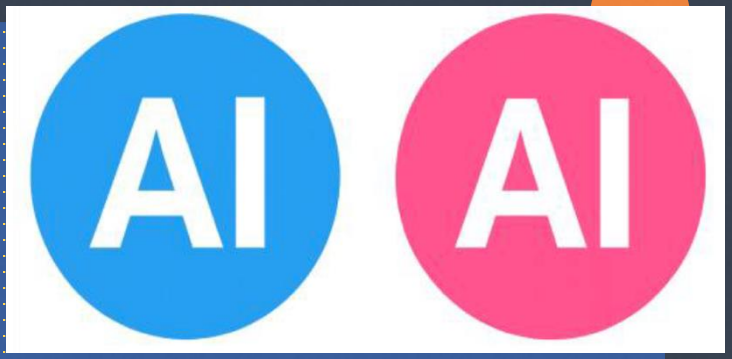


Figure 1. A zero-shot prompt on ChatGPT's free version as of December 2025 for an icon containing the word AI in two different colours indicating the difference between fully automated and AI-assisted content.

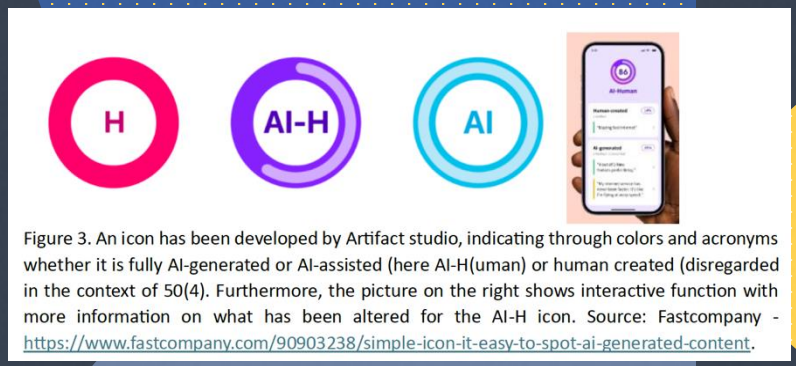


Figure 3. An icon has been developed by Artifact studio, indicating through colors and acronyms whether it is fully AI-generated or AI-assisted (here AI-H(uman) or human created (disregarded in the context of 50(4). Furthermore, the picture on the right shows interactive function with more information on what has been altered for the AI-H icon. Source: Fastcompany - https://www.fastcompany.com/90903238/simple-icon-it-easy-to-spot-ai-generated-content.

Prepared by T.Jitsuzumi for 1/6/2026.

Section 2（AIシステムデプロイヤー向けルール）

Commitment 4: Specific Measures for Deepfake Disclosure

- Measure 4.1：ディープフェイク判定の内部プロセス確立＋判定作業への人間の関与の確保
- Measure 4.2：ディープフェイク初回露出時までには以下の様式で情報表示
  - Sub-measure 4.2.1：リアルタイム動画＝可能な範囲でアイコン常時表示＋冒頭ディスクレーマー
  - Sub-measure 4.2.2：非リアルタイム動画＝アイコン（表示タイミングは選択可能）
  - Sub-measure 4.2.3：その他マルチモーダルコンテント＝アイコン常時表示など
  - Sub-measure 4.2.4：画像＝露出ごと固定位置に表示
  - Sub-measure 4.2.5：音声＝短尺（30秒未満）は冒頭に音声でディスクロージャー挿入、長尺は複数回ディスクロージャー挿入 ＋画面あればアイコン
- Measure 4.3：芸術創作物等は鑑賞を阻害しない“非侵襲”表示＋第三者の権利配慮

Commitment 5: Specific Measures for Disclosure of AI-Generated and Manipulated Text

- Measure 5.1：法50条(4)に該当するか否かの内部判定プロセスの確立
- Measure 5.2：初回露出時までに固定位置でアイコン表示（上部/横/奥付/末尾など）
- Measure 5.3：事業者規模に応じた法50条(4)の例外に該当するか否かの内部判定プロセス等の確立

Prepared by T.Jitsuzumi for 1/6/2026.

パブリックコメント

1. それぞれの措置（Measure, Sub-measure）についての賛否とその理由
    - この措置は必要な水準に近い
    - この措置は軽微な修正および／またはさらなる明確化が必要
    - この措置は大幅な修正および／またはさらなる明確化が必要
    - この措置は完全に削除すべきである

2. 次ページ以下の自由回答設問
    1. AIシステムプロバーダー向け4問
    2. AIシステムデプロイヤー向け5問

自由回答設問１

AIシステムプロバイダー向けルールについて

1. 電子透かしの技術的実現可能性
   - 電子透かしを適用する際に考慮すべき、特定のコンテンツタイプやその長さ/形式に関連する制限はありますか？また、特定のケース（コードや画像など）において、特定の閾値が関連するか不要かについても含めてください。
2. 新興AIコンテンツタイプに対する追加措置
   - ソフトウェアコード、エージェント型AI、ゲーム、VR、音声アシスタント、および第1草案で対象外となったその他の新規AI生成コンテンツに対して、どのような措置やコミットメントが適用されるべきと考えますか？
3. AI生成コンテンツの共通検証システム
   - 経済的・安全保障上の制約を考慮し、AI生成コンテンツの共通検証システム（複数のAIシステム／モデル提供元が発行するマーキングを検出・検証する仕組み）を技術的に実装するにはどうすべきか？ こうした検証システムは中央集権型と分散型、どちらが適切ですか？
4. その他の技術的提言
   - 公的インフラ・サービスに関して、他に提起すべき技術的考慮事項や提言はありますか？

第一草案で対象となっているAI生成コンテンツ

the text, image, video or audio content, or any combination thereof, generated or manipulated by the AI system(s)

自由回答設問２

AIシステムデプロイヤー向けルールについて

1. Measurement 1.1関連
    - 説明されている分類体系に反映されていない具体的な事例や例を提案してください。
2. Measurement 1.2関連
    - 提案されている要件と同様の要件を満たす、既存の展開済みラベルやアイコンへの参照を提供してください。分類体系を反映するのに必要な細かさを持ち、異なるモダリティやインターフェース（画面/音声コンテキストを含む）間で適応可能なアイコンの有力な事例を求めています。適用済みで明確かつ識別可能なAI関連アイコンの事例（共通参照基準となり得るもの）は、CoP第2草案の開発において極めて有用です。
3. Commitment 1関連
    - 分類体系記述、暫定アイコン、EUインタラクティブアイコンのさらなる発展を支援できる関連学術研究の参照先を提供してください。特に、ラベルやアイコンが自然人に対してコンテンツのステータスを伝える有用性を検証した研究（例：広告ラベルによるコンテンツ識別に関する研究、視線追跡研究、ユーザーが時間経過とともにアイコンを認識するか否かの検証、インタラクティブアイコンと非インタラクティブアイコンの評価）に関心があります。

自由回答設問３

AIシステムデプロイヤー向けルールについて

4. Measurement 2.3関連
    - 措置2.3の具体化に役立ち、かつフラグ付けサービスの促進という目的を達成すると考えられる既存技術や実装例に関する参考文献をお持ちですか？
5. Measurement 4.1関連
    - 署名者が内部プロセスを策定する際に考慮すべき関連対象者を示していただけますか？未成年者や障がい者などのカテゴリーは明らかに対象範囲に含まれますが、特定のニーズや脆弱性が考慮されるべき追加グループも特定したいと考えています