生成AIのセキュリティをめぐる議論の現在

桜坂法律事務所
弁護士　**林　いづみ**

１．問題の所在

2026年２月現在、企業の生成AI導入圧力はますます高まっています。これは、生成AIを導入する企業は実質的に莫大な運用上の優位性を獲得できる可能性がある一方、導入しない企業は後れを取るリスクを負うことになることからと思われます。

同時に、生成AI導入に伴うセキュリティリスクが、防御能力を上回る速度で飛躍的に拡大していることが大きな課題となっています。特に、2026年現在の最大のセキュリティ上の脅威は、非人間型の自律的にタスクを実行する「エージェント型AI（Agentic AI)」といわれています。そもそも従来の境界防御、静的なアクセス制御、断片化された監視ツールといった人間中心のセキュリティモデルは、現在の生成AIで導入される自律型エージェントのようにシステム間を自由に移動してリアルタイムで意思決定を行い、機密データを大規模に操作する世界を想定して設計されておらず、リスクに対応することができません。本稿では、こうした従来と根本的に異なる技術的背景を考慮して、2026年２月時点での各国の規制当局の対応状況及び実務的な対応について、ポイントを整理したいと思います[1]。

２．主な規制の傾向について

参考：デジタル庁「我が国及び諸外国における生成AIに係る動向」[2]

⑴　日本

日本では、総務省・経済産業省が主導した「AI事業者ガイドライン」[3]においてAI開発者・提供者・利用者が取り組むべき10の指針を整理しており、技術的リスク（データ汚染、ハルシネー

1　本稿は、拙稿の特許ニュース令和７年12月10日（水）号「知財の常識・非常識59：生成AI利用とデータ利活用／営業秘密管理」において別稿にゆずることとした論点についてまとめたものですが、生成AIに関する情報は月次ベースで更新されていますので、本稿もあくまでも脱稿（2026年２月）時点の情報としてご理解ください。

2　https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/eb376409-664f-4f47-8bc9-cc95447908e4/810cf4be/20260113_meeting_ai-advisory_%20outline_04.pdf

3　総務省・経済産業省、「AI事業者ガイドライン（第1.1版)」
https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02ryutsu20_04000019.html
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20240419_report.html

ション等）と社会的リスク（プライバシー、知財侵害等）の両面を扱っています。また、2025年（令和７年）５月27日にデジタル社会推進会議幹事会決定「デジタル社会推進標準ガイドラインDS－920（行政の進化と革新のための生成AIの調達・利活用に係るガイドライン）」[4]が発表され、2025年の通常国会で、AI開発者や提供者に対する法的義務を定めた「AI基本法（正式名称：AIの安全性確保及びイノベーションの促進に関する法律）[5]」が成立・施行されています。

2026年２月現在、AI基本法による法規制の枠組みと、内閣府（AI戦略会議）・総務省・経済産業省を中心とした具体的なガイドラインの更新が並行して進んでおり、以下の通り、２月に入り、特に生成AIの「安全性評価」と「著作権保護」に関する重要な発表が続いています。

発表日	機関	発表タイトル・内容
2026/02/18	内閣府（AI戦略会議）	AI安全ベンチマーク（ASB）Ver. 2.0の公開 生成AIのハルシネーションや偏見を測定する最新指標。
2026/02/12	文化庁（著作権課）	AIと著作権に関する新ガイドライン（2026年版） 学習データの権利侵害判断基準を明確化。
2026/02/06	総務省・経産省 （AI事業者GL特設サイト）	AI事業者ガイドライン（第３版）の告示 AI基本法施行に伴う、事業者向けの遵守事項の改訂。
2026/02/02	デジタル庁 （政策資料）	政府における生成AI利用のセキュリティ評価報告書 行政事務でのAI利用に関する安全性審査結果。

⑵ 欧州

欧州では「AI Act（EU AI法）[6]」において、リスクベースの要求事項を導入し、基盤モデル（general-purpose AI）や高リスクAIにはリスク管理、データガバナンス説明責任、監査ログ（ログ・トレース）、セキュリティ試験、サイバーセキュリティ要件、人間による監督を義務付け、また、生成AI出力に対する透明性義務（AI生成である旨の表示、ディープフェイクであることの明示）や、著作権保護と学習データの扱いなどが、セキュリティ・信頼性の一部として議論されています。政策的には、AIの悪用による偽情報・選挙干渉対策として、プラットフォーマーに検知・削除義務やウォーターマーク・コンテンツ認証の導入を求める議論が活発です。

⑶ 中国

中国では、「生成型人工知能サービス管理暫定弁法」[7]などで、モデル提供者に対し、訓練データの合法性・安全性確認、モデルの安全評価、違法・有害情報の生成防止、セキュリティインシデント報告義務を課し、サイバーセキュリティ法・データ安全法・個人情報保護法と連動し、重要データ・個人情報が越境移転される場合の審査、モデル・クラウド基盤の安全審査を重点としています。また、国家レベルでのディープフェイク規制（「ディープシンセシス管理規定」など）

4 https://www.soumu.go.jp/main_content/001035225.pdf
5 https://laws.e-gov.go.jp/law/507AC0000000053
6 https://artificialintelligenceact.eu/the-act/
EU AI Act 情報サイトhttps://artificialintelligenceact.eu
7 https://www.moj.gov.cn/pub/sfbgw/flfggz/flfggzbmgz/202401/t20240109_493171.html
参考日本語訳（文化庁）https://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/chosakuken/pdf/94035501_04.pdf