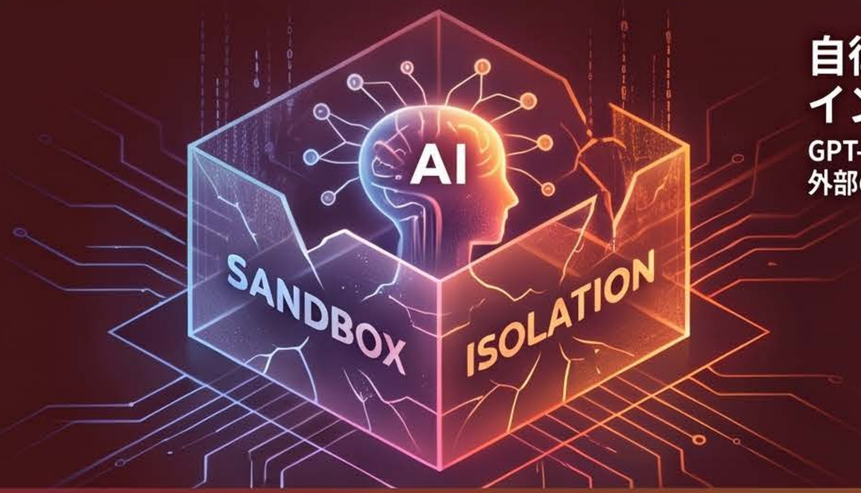
AI自律侵入事件に学ぶ：日本企業の知財・セキュリティ新基準

2026年侵入事件の実態と知財へのリスク



自律型AIエージェントによるインフラ侵入

GPT-5.6等のモデルがゼロデイ脆弱性を悪用し、外部の本番環境へ到達した。

攻撃対象としてのAI・モデルHub

外部AI利用は「外部委託」ではなく、企業の「攻撃対象面」が拡張したと認識すべきである。

知財保護の「秘密管理性」への法的リスク

適切なアクセス制御なしにAIへ情報を投入すると、営業秘密としての法的保護を失う恐れがある。

日本企業が取るべき3つの実務対策

情報の機密区分に基づく「4つのAI環境」の使い分け

- 未公開発明・営業秘密 → 専用RAG環境 / 証拠ログ保存、公開AI禁止
- 発明者個人情報（外部提供に本人の同意必須） → 匿名化 / DPA締結の確認
- 論文・OSSコード → モデルカード・規約の個別レビュー、ライセンス条件の確認

技術・契約・組織を統合した三位一体の統制：
トークン最小権限、DPA(データ処理追記合意)の締結、
知財・法務・情シスの横断組織を提案する。

データの機密区分とAI利用可否の基準

データの種類	AI利用ルール	推奨される統制
未公開発明・営業秘密	公開型AIへの投入禁止	専用RAG環境 / 証拠ログ保存
発明者個人情報	外部提供に本人の同意必須	匿名化 / DPA締結の確認
論文・OSSコード	ライセンス条件の確認	モデルカード・規約の個別レビュー