期待通り、予想以上。PTW

AIガイドライン
何から始めればいいのか

OWASP・NIST・EU AI法・ISO 42001国際フレームワークを羅針盤にする

JaSST Kansai 2026　セッション

ポールトゥウィン株式会社
事業本部 QAソリューション事業部
事業推進グループ

AIの活用状況と課題

① 活用は広がっている

34.5%

の企業が生成AIを活用

活用企業の 86.7% が「業務効果あり」と回答

出典：帝国データバンク「生成AIに関する企業の動向調査」

② しかし、課題も多い

1位	情報の正確性	50.4%
2位	専門人材・ノウハウ不足	41.3%
3位	活用すべき業務の範囲	40.0%
4位	情報漏洩のリスク	33.5%
5位	責任所在・ルール整備	25.5%

出典：帝国データバンク「生成AIに関する企業の動向調査」

AI活用は進めなければならないが、活用には「ガイドライン整備」で課題の対処を行う必要あり

AIガイドラインの現状

全体の策定率

スリスタ調査（2026年5月）

11.5%

形骸化含めても 18.3%
中小企業(1-50名)：0%
大企業(5001名以上)：28.8%

AI活用企業に限定

総務省データ分析（2026年3月）

約50%

使っている企業の半数が
ルールなし「野良AI状態」

大手企業の管理職層

パーソルキャリア（2025年9月）

4〜5割

AI導入率は約8割
成果企業ほどルール整備と
人材育成をセットで実施

→ あなたの組織に問いかけてみてください

- Q1 AIを導入したとき、何かトラブルが起きたら誰が責任を取りますか？
- Q2 社員がAIに入力してよい情報・してはいけない情報、明文化されていますか？
- Q3 AIの出力をそのまま使っていませんか？検証プロセスはありますか？

「即答できない」— それがガイドライン整備の出発点です

事例紹介：PTWではどのようにしてガイドラインを策定したか

ガイドライン策定の背景

国内外で顕在化したリスク

・IPA「情報セキュリティ 10 大脅威 2025」で生成 AI 関連が **初選出・第 3 位**
・AI エージェント特有のインシデント事例（機密情報流出・不可逆操作）が主要事業者から **複数公表**
・国内 AI 推進法（2025-06 公布）で **規制対応が必須化**

従来統制では捕捉できない理由

・**自律接続**
ヒトの介在なく外部 API を実行する
・**自律実行**
ヒトの承認なく不可逆操作（削除・送信・決済）を完了する
・**入力経路の拡大**
画像・ファイル等の経路からプロンプトインジェクションが成立する

取りうる対応と不作為リスク

・**統制なしの利用拡大**
インシデント時の責任主体が不明になる
・**全面禁止**
競合優位の喪失とシャドー IT 化を誘発する

PTWでは下記の３ステップでガイドラインの策定を実施

1 **自社のAI利用シーンを棚卸しする**
今どんな業務でAIを使っているか、どのようなAIモデルを使用しているか書き出す
例：文書作成、コード補助、顧客対応チャット、採用スクリーニング…
→ まず「見える化」から

2 **リスク分類で優先度をつける**
棚卸ししたシーン、ツールをリスクレベルに当てはめる
高リスクに該当するものから優先的に対応を検討する
→ リスクを数値化

3 **具体的な対策を設計する**
対策の有効性はフレームワークに照らし合わせて評価する
問題がなければ、ガイドラインとして文書化する
→ 国際規格を活用

事例紹介：PTWではどのようにしてガイドラインを策定したか

PTWでは既存ISMSの延長線上にAIエージェント固有の論点を上積みし、**情報セキュリティの維持**と**AI活用による生産性向上**、**社員の自律的成長**を実現させるためのAIガバナンスを策定。

ISMS(ISO/IEC 27001)を基盤にして、
AIエージェント固有の論点として、右記の４観点５文書を取り込んでAIガバナンスを策定

取り込み観点	取り込んだ企画	制度に採用した要素
AIマネジメント体系の骨格	ISO/IEC 42001:2023 NIST AI RMF 1.0(生成AI対応含む)	AIMSのPDCAサイクル・管理策・リスク管理プロセス
AI固有の脅威・脆弱性	OWASP LLM Top 10(2025)	LLM固有脅威10項目(プロンプトインジェクション・過度のエージェンシー・サプライチェーン攻撃等)への統制
国内規制対応	経産省AI事業者GL v1.1	10原則・アジャイルガバナンス・AI事業者GL/AIガバナンスGLの主要要求事項
ユースケース分類の軸	EU AI Act(Regulaton 2024/1689)	リスクベース４段階のユースケース判断基準(禁止/高リスク/限定/最小)

事例紹介：PTWではどのようにしてガイドラインを策定したか

リスクアセスメントの実施

AI利用形態を自律性の度合いと主たる操作インターフェースで３層に分類。レイヤーごとに４大分類24項目のリスクアセスメントを実施。

分類A　外部接続（6項目）

何に繋ぐか

MCP／WebSearch／Git 等

分類B　ローカル（6項目）

何に触るか

ファイル／コマンド／Computer Use 等

分類C　制御（7項目）

いつ・どう動くか

Hooks／Cron／Channels 等

分類D　注入（5項目）

何者にするか

SKILL.md／CLAUDE.md／Projects 等

レイヤー	種別	自律実行可	リスク	利用者層
Layer 1	対話型 AI	3 / 24	低	
Layer 2	自律型 AI（GUI）	14 / 24	高	
Layer 3	自律型 AI（CLI）	22 / 24	最高	

利用者層については、次ページで説明する４ガバナンス原則によりコントロールする

事例紹介：PTWではどのようにしてガイドラインを策定したか

4つの統制原則による多層防御の設計（仕組み③④と運用①②でリスクを制御）

原則 1
ホワイトリスト運用
承認済みツール・接続先のみ利用
9 項目カバー

原則 2
Human-in-the-loop
不可逆操作には人間の承認
8 項目カバー

原則 3
最小権限・サンドボックス
AI 専用作業領域・管理者権限禁止
12 項目カバー

原則 4
外部チャネル遮断
外部メッセンジャー等の遠隔操作禁止
4 項目カバー

4つの統制原則による多層防御が有効であることを国際フレームワークを用いて評価

4 つの統制原則

- ホワイトリスト運用
- Human-in-the-loop
- 最小権限・サンドボックス
- 外部チャネル遮断

主要なフレームワークへの対応

- ISO 42001
- NIST AI RMF
- EU AI Act
- 経産省ガイドライン
- OWASP LLM Top 10

主要なフレームワークの要求事項と整合を確認することにより、4つの統制原則による多層防御で、24項目のリスク対応に有効であることを確認

OWASP Top 10 for LLM

概要	OWASP Top 10 for LLM Applications は、大規模言語モデル（LLM）を活用したアプリケーションに固有のセキュリティリスクを体系的に整理し、開発者・セキュリティ担当者・意思決定者に対して優先的に対処すべき脅威を提示しているセキュリティ啓発ドキュメント。

原則１：ホワイトリスト運用

LLM03	**サプライチェーン** サードパーティコンポーネント管理
LLM06	**過度のエージェンシー** 機能の最小化

原則２：Human-in-the-loop

LLM06	**過度のエージェンシー** 機能の最小化
LLM10	**無制限消費** リソース消費の制御

原則３：最小権限・サンドボックス化

LLM01	**プロンプトインジェクション** 入力のサンドボックス処理
LLM02	機密情報漏洩 アクセス範囲の限定
LLM06	**過度のエージェンシー** 機能の最小化

原則４：外部遠隔チャネルの原則遮断

LLM01	**プロンプトインジェクション** 入力のサンドボックス処理
LLM10	**機密情報漏洩** アクセス範囲の限定

NIST AI RMF

概要	NIST AI Risk Management Framework（AI RMF）は、AIシステムの設計・開発・利用・評価において Trustworthiness（信頼性）の考慮事項を組み込む能力を向上させるために開発されたフレームワーク。 AI RMF Coreという４つの機能で構成されている

GOVERN	MAP	MEASURE	MANAGE
AI方針・責任体制・判断基準を定める	リスクを特定・分類する	リスクを測定・評価する	対策を実施・継続的に改善する

原則１：ホワイトリスト運用

MAP 3.4	オペレーター習熟度に関するプロセス定義
MANAGE 2.2	本運用AIシステムの価値維持メカニズム 承認済みツール構成のドリフト防止

原則２：Human-in-the-loop

GOVERN 1.4	リスク管理プロセスと成果の透明な方針の確立
MEASURE 2.6	安全リスクの定期評価と安全性メトリクス
MANAGE 2.4	パフォーマンスまたは意図外の結果に対する無効化・廃止手順 人間承認による停止判断

原則３：最小権限・サンドボックス化

MAP 1.6	AIシステムの境界の定義
MANAGE 1.3	高優先度リスクへの対応計画

原則４：外部遠隔チャネルの原則遮断

GOVERN 6.1	外部リスクの管理
MANAGE 4.2	継続改善活動の統合と関係者との定期関与

ISO/IEC 42001

概要	ISO/IEC 42001は、AIに特化した世界初のマネジメントシステム国際規格。組織がAIシステムを責任を持って開発・提供・利用する為の方針・プロセス・管理策を体系的に定め、PDCAサイクルによる継続的改善を実現する為の枠組みを提示している。ISO 27001と構造的に高い親和性を持っている為、統合的な運用が推奨される。

原則１：ホワイトリスト運用

A.6	**AIシステムライフサイクル管理** ツール・コンポーネントの選定管理
A.10	**第三者関係** サードパーティAIコンポーネントの管理

原則２：Human-in-the-loop

A.4	**AIシステムのためのリソース** 人間による監視
Clause 8	**運用** プロセスの計画と管理

原則３：最小権限・サンドボックス化

A.6	**AIシステムライフサイクル管理** アクセス制御
ISO27001統合	**管理策A.8** アクセス制御 **管理策A.5.15** アクセスの権限管理

原則４：外部遠隔チャネルの原則遮断

A.10	**第三者関係** 外部接続の管理
ISO27001統合	**管理策A.5.19** サプライヤー管理 **管理策A.8.20** ネットワークセキュリティ

EU AI法

概要	EU AI Act（Regulation(EU) 2024/1689）は、世界初の包括的なAI規制法。EU域内市場の適切な機能を確保しつつ、人間中心で信頼できるAIの普及を促進し、健康・安全・基本的権利（民主主義、法の支配、環境保護を含む）をAIシステムの有害な影響から高い水準で保護することを目的として策定されている。

容認不可	高リスク	限定リスク	最小リスク
社会的スコアリング・潜在意識操作	採用・与信・医療診断・重要インフラ	チャットボット・AIコンテンツ生成	スパムフィルター・レコメンド
→全面禁止	→厳格な義務	→透明性義務	→原則自由

原則１：ホワイトリスト運用

Art.15	**正確性・ロバスト性・サイバーセキュリティ** サプライチェーン管理
Art.9	**リスク管理システム** 適切な安全装置

原則２：Human-in-the-loop

Art.14	**人間による監視** 高リスクAIシステムに対する人間の監視義務

原則３：最小権限・サンドボックス化

Art.9	**リスク管理** 適切な安全装置
Art.15	**サイバーセキュリティ** アクセス制御

原則４：外部遠隔チャネルの原則遮断

Art.9	**リスク管理** 外部攻撃面の最小化
Art.13	**透明性** 遠隔操作の開示

経産省AI事業者ガイドライン

概要

「AI事業者ガイドライン」は、AIガバナンスの統一的な指針としてイノベーションの促進とリスクの緩和を目的として策定。元々あった「AI原則実践のためのガバナンス・ガイドライン」「AI利活用ガイドライン」「AI・データの利用に関する契約ガイドライン」を統合して策定された。

人間の尊厳を守る社会	多様な幸せを包摂する社会	地球規模で持続可能な社会
AIが人間の能力を拡張し、憲法や国際人権規範を保障しながら、多様な人々の幸福追求を可能にする	バイアス対策と包摂性を確保し、多様な価値観を尊重する社会の実現	環境負荷の低減や持続可能な発展に寄与するAIの活用

原則１：ホワイトリスト運用

原則5	**セキュリティ確保** AIシステムの安全性確保
原則7	**アカウンタビリティ** 利用ツールの説明責任

原則２：Human-in-the-loop

原則1	**人間中心** 人間の判断の介在

原則３：最小権限・サンドボックス化

原則5	**セキュリティ確保** 最小権限の原則
原則4	**プライバシー保護** データアクセスの制限

原則４：外部遠隔チャネルの原則遮断

原則5	**セキュリティ確保** 不正アクセスの防止
原則4	**プライバシー保護** 外部からのデータ窃取防止

まとめ

1. AIの活用は広がっているが、約半数の企業がガイドラインを整備できていない
2. 課題（正確性・漏洩・責任・人材・業務範囲）には対応するフレームワークがある
3. OWASP・NIST・EU AI法・ISO 42001・AI事業者ガイドラインは「羅針盤」として今日から使える
4. まずはAIの見える化から行ってみよう

今回お話しさせていただいた内容は、PTW内の一例に過ぎませんが、
このセッションが、皆さんの組織にとっての羅針盤を見つけるきっかけになれば嬉しいです。

参考資料

フレームワーク

- OWASP Top 10 for LLM
- NIST AI RMF
- EU AI Act
- ISO/IEC 42001

国内ガイドライン

- 経産省・総務省「AI事業者ガイドライン v1.2」

調査データ

- 帝国データバンク「生成AIに関する企業の動向調査」
- 株式会社スリスタ「AIガイドライン策定実態調査（2026年5月）」
- パーソルキャリア HiPro Biz「大手企業役職者AI意識調査（2025年9月）」

期待通り、
予想以上。PTW