

Gemini 3.5 Flash 「Computer Use」が変える知財実務：RPAを超えたAIエージェントの衝撃

Context Summary

GoogleのGemini 3.5 Flash 「Computer Use」は、AIが人間のようにPC画面を視覚認識し、自律操作を可能にする。従来のRPAの「画面デザイン変更で停止」する脆弱性を克服。APIのないレガシーシステムや複数国の特許庁ポータルを跨ぐ知財業務（情報転記、文献取得、品質チェック等）の高度な知識労働を自動化。機密情報保護のためZDR設定や人間介在が不可欠。

従来型RPA

UI変更への耐性の劇的な向上：
ピクセル位置やDOM構造に依存、数ピクセルずれで停止（低耐性）
単純な定型作業から「推論を伴う知識労働」へ：
データの転記など定型的繰り返し作業のみ
開発・保守コストの削減：
模範シナリオ構築やレコーディングが必要

VS.

Gemini 3.5 Flash (Comini 3.5 Flash)

UI変更への耐性の劇的な向上：
視覚情報から「これがログインボタン」と意味的に推論し継続（極めて高い耐性）
単純な定型作業から「推論を伴う知識労働」へ：PDF要約、ステータスの動的変更など非定型・推論を伴う作業に対応
自然言語（プロンプト）による指示だけでAIエージェントにタスク実行

	従来型RPA	比較項目	従来型RPA	Gemini 3.5 Flash (Computer Use)
UI変更への耐性	低		低	高い
対応可能なタスク	ルーティンタスク		ルーティンタスク	非定型・推論を伴う作業
例外処理・回復	停止		ストップ	自律的に回避・回復
インターフェース	シナリオ構築ツール		シナリオツール	自然言語

AIが画面を操作する仕組み：エージェントループ

1. 環境状態の送信と視覚認識

ユーザー指示と現在の画面スクリーンショットをGeminiに送信



2. 推論と意図 (Intent) の生成

画面解析、次の操作を決定、操作の「意図」も出力（透明性確保）



4. 再評価と反復

操作後の新画面を再キャプチャしフィードバック、タスク完了までループ



3. 座標の非正規化とアクション実行

1000x1000グリッドで認識した操作位置を実際の解像度に合わせクリック/タイピング実行



知財実務における3つの革新的ユースケース



クライアントポータルと所内DBの自律連携

顧客別ポータルから新規案件データ抽出し、APIのない所内特許管理システムへ正確に転記・登録



引用文献の自動取得と知的なファイル命名

拒絶理由通知から引剤番号特定、J-PlatPat等からPDFダウンロード、内容解析し事務所規則でリネーム・保存



明細書の継続的品质監査（符号照合等）

Wordからテキストをコピー、外部符号照合ツールで実行、結果を読み取りレポートをWordに追記するGUI操作を代行



安全な導入のための「多層防御」とコンプライアンス

Zero Data Retention (ZDR) の徹底



未公開の発明情報



Vertex AI環境



(1) データキャッシングを無効化



(2) 乱用監視ロギングのオプトアウトを申請



ZDR達成（データ保持ゼロ）

Human-in-the-loop（人間の関与）の設計



特許庁への出願、納付期限の書き換えなど不可逆で重大な操作前はAIを一時停止、人間の承認を要求



グラウンディング機能の回避

Google検索等によるグラウンディング機能はデータが30日間保存されるため、極秘タスクでは利用回避