資料 1

Graffer

AI革命で仕事はこう変わる

株式会社グラファー 代表取締役 石井 大地

変貌するソフトウェア開発プロセス

AI革命のさらなる進展

- この1年でAI技術はさらに進展し、ユースケースが爆発的に広がった

1 Agentic AI（AIエージェント）の一般化

単に指示や質問に応答するAIチャットとしての使い方よりも、AIが人間の意図をくみ取って自律的に作業計画を立て、時間をかけながらタスクをこなしていく「Agentic AI」が当たり前に使われるようになった。
一般的なAIチャットサービスも、推論モデルがウェブ検索を駆使してタスクをこなすようになっている。

2 マルチモーダル性能の劇的な進化

画像、音声、動画の入出力性能が大幅に向上したことで、テキストベースではなく音声会話でAIとやりとりすることも実務上問題のないレベルで可能になった。
またエクセルやパワーポイント、PDFなど業務で扱うファイルの認識・操作能力も著しく向上している。

3 オープンウェイト・オープンソースの台頭

ローカルのコンピューターやサーバーで動かせる、オープンウェイト・オープンソースのモデルについても性能が大きく向上した結果、安価で、かつ安全な環境でAIを活用できる用意が整った。
センシティブな個人情報・機密情報を扱うケースでも利用できる場所が急速に広がりつつある。

AIの能力のうち、プログラミング性能の向上は特に著しい

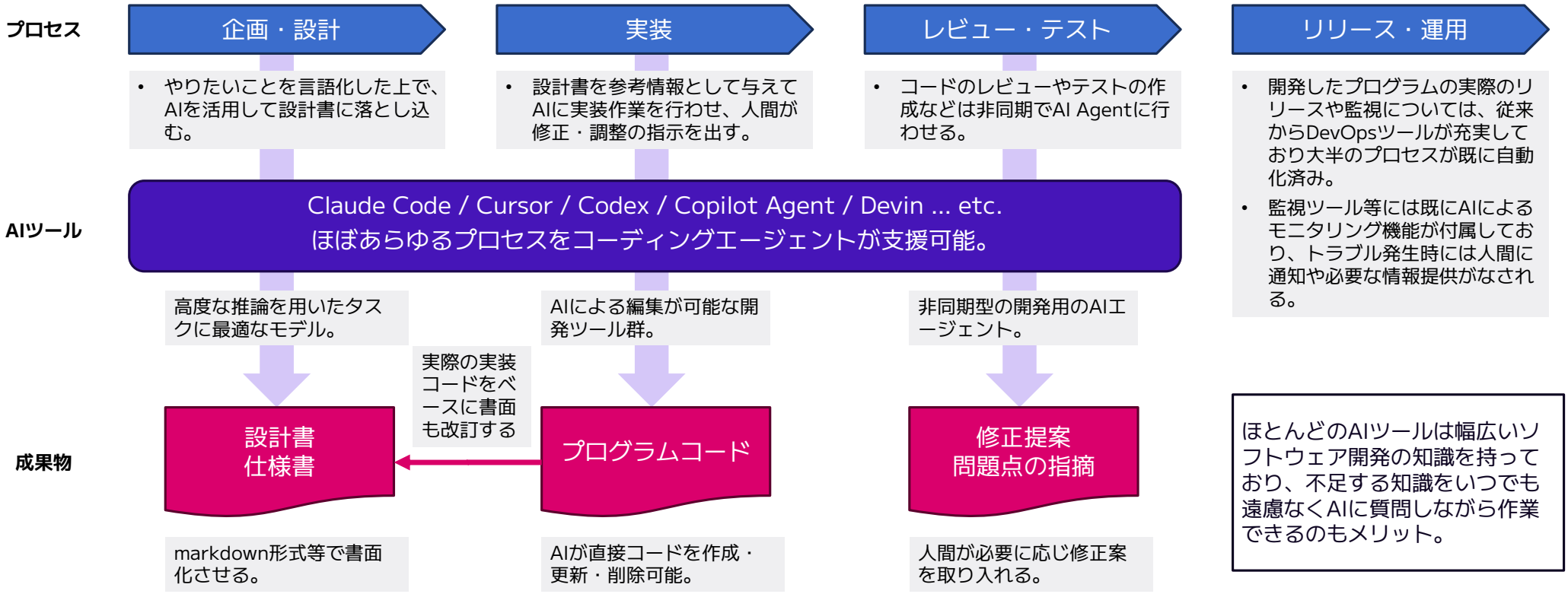
DeepSWE v1.1 (Datacurveによる共通条件での評価, *は各社公表ベース)

登場年月	モデル名	スコア
2026-5	Claude Opus 4.8	59.0%
2026-6	Claude Fable 5	69.9%
2026-7	GPT-5.6 Sol	72.7%
2026-7	Gemini 3.6 Flash	46.7%
2026-7	Claude Opus 5	73.7%
2026-8	Gemini 3.7 Flash	65.5%
2026-9	Claude Fable 5.1	67.4%
2026-9	Gemini 3.8 Flash	73.8%
2026-9	**GPT-6 Astra**	**74.1%**
2026-9	**Claude Opus 5.5**	**74.2% ***
2026-9	**GPT-6.1 Sol**	**75.2% ***
2026-9	**Gemini 4 Argon**	**77.9% ***

- ✓ 様々なベンチマークテストでのソフトウェア開発の性能は、時間とともに急速に改善している事実がある。
- ✓ 様々なベンチマークが用意されては1年も経たずにスコアが上限近くに張り付いて無意味化することを繰り返している。
- ✓ 特に、2026年9月のモデル進化は過去3年を見てきたなかでも異常なレベル。

AI革命で変わるソフトウェア開発プロセス

- AIの存在を前提とすることでワークフローが完全に変貌し、従来なら数名～数十名で実施していた業務のほとんどを1名で自己完結的に実施可能な時代になった。



AIを活用したソフトウェアエンジニアリングの重点1

- AIの使い方も進化し、プロンプトやコンテクストの制御にこだわる手法は陳腐化し、ハーネスやループの設計によってAIエージェントによる自律実行の仕組みづくりに焦点が変わってきている。

2023～2024	2024～2025	2025～2026	2026～
プロンプトエンジニアリング	**コンテクストエンジニアリング**	**ハーネスエンジニアリング**	**ループエンジニアリング**
プロンプトの書き方を工夫することでAIとのチャットからより良いアウトプットを導こうとする手法。	AIエージェントがタスクを実行する際に依拠できるコンテクストをあらかじめ用意して挙動を制御する手法。	AIエージェントがタスクを安全かつ有効に実行できる環境・設定を整備するための手法。	AIが必要なプロセスの改善を行いながらタスクを実行し続けるループを作り生産性を最大化していく手法。
【陳腐化の背景】 AIの性能向上により、細かいプロンプトを工夫する必要はなくなり、むしろ技巧的なプロンプトにより性能が低下する局面も出てくるようになった。	**【陳腐化の背景】** 大量のコンテクストを与えることによる”Lost in the middle” や過大なトークン消費の課題や、AIエージェント利用にかかるセキュリティ対策が課題に。	【利用シーン】 • コーディングエージェントによるタスクのクラウド実行 • 専用の隔離されたマシン上でのOpenClawの実行	【利用シーン】 • システムにおける自己改善メカニズムの実装 • AI CEO framework / Software-defined Company
コンテクストエンジニアリングへ	ハーネスエンジニアリングへ	**ソフトウェア開発手法の最前線も、 現状では既にハーネスの整備やループ設計に移行している。**	

AIを活用したソフトウェアエンジニアリングの重点2

- ハーネスエンジニアリングやループエンジニアリングの次は、複数のAIエージェントやその他の処理を組み合わせて動かすグラフエンジニアリングの考え方も広がってきた。

2025～2026	2026	2026後半
ハーネス エンジニアリング	**ループ エンジニアリング**	**グラフ エンジニアリング**
AIエージェントがタスクを安全かつ有効に実行できる環境・設定を整備するための手法。	AIが必要なプロセスの改善を行いながらタスクを実行し続けるループを作り生産性を最大化していく手法。	複数のAIエージェントやその他の処理を組み合わせ、業務を実行するためのアーキテクチャを作り込む手法。
【利用シーン】 • コーディングエージェントによるタスクのクラウド実行 • 専用の隔離されたマシン上でのOpenClawの実行	【利用シーン】 • システムにおける自己改善メカニズムの実装 • AI CEO framework / Software-defined Company	【利用シーン】 • 複数のエージェントによって業務を遂行するアーキテクチャの実装 • LangGraph, *"Graph Engineering in the Era of LLM Agents"*

ソフトウェア開発手法の最前線も、
現状では既にハーネスの整備やループ設計に移行している。

ハーネスやループの整備を前提として
複雑なタスクを処理する構造が重要。

Key Changes:

- ✓ ソフトウェアの開発は、エンジニア組織によって人力で行われる業務ではなくなった。
- ✓ AIを用いて稼働する **Software Factory**が業務の中心と位置付けられるようになった。
- ✓ 開発者の業務はコードを書くことではなくSoftware Factoryの生産性と品質を高める工程の改善が軸となっている。

ソフトウェアの業界で起きたことが、今後すべての業界で起こる

- AIの能力は汎用的でない
  - 様々な研究を包括すると、AIの知的能力はまんべんなくすべて高いのではなく、ソフトウェア開発などいくつかの領域においてずば抜けて高いなど、能力に大きなばらつきがある状態である。
- ……しかし、プログラミング能力があれば、**大抵のことはできてしまう**
  - AIのソフトウェア開発能力が劇的に高まった結果、動画やプレゼンスライド、ミュージックビデオなども、画像生成や動画生成のモデルを使うのではなく、プログラムを使って作る方法が台頭している。
  - 例えば、AIに動画生成をさせても完ぺきに狙い通りの動画を作るのは難しいが、3DCGのセットをAIに作らせ、その上でプログラムを走らせてカメラや3Dモデルを動かせば狙い通りの挙動を細かく制御できる。
  - このトレンドが続けば、AIによる業務代替は、「AIによるソフトウェア開発」を経由して行われる可能性が高いのではないか。

ループエンジニアリングと人間の役割

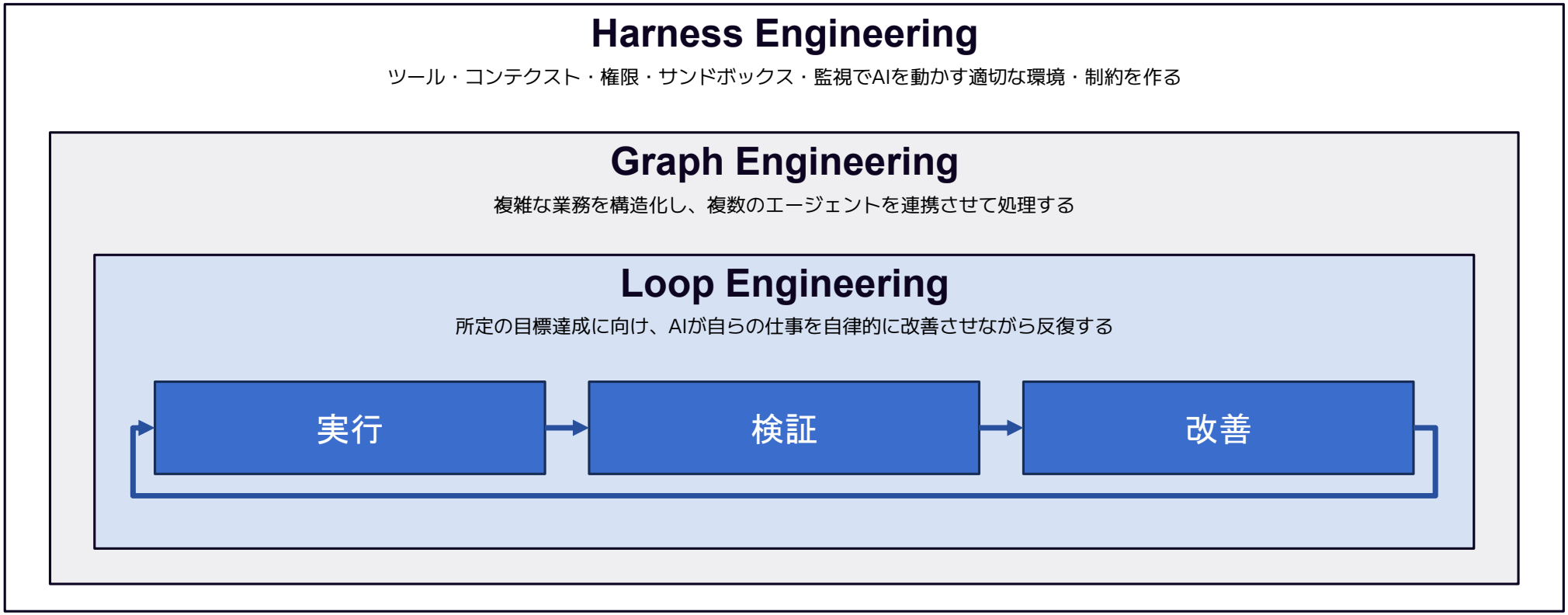
認知タスクにおいてAIがほとんどの人間を上回る前提の業務分担とは

- 業務の設計や管理すらAIを使いながら組み立てていくのが当たり前になり、人間はAIの行動を方向付けたり、許容できるラインをガードレールとして作り込むことが役割に変わる。



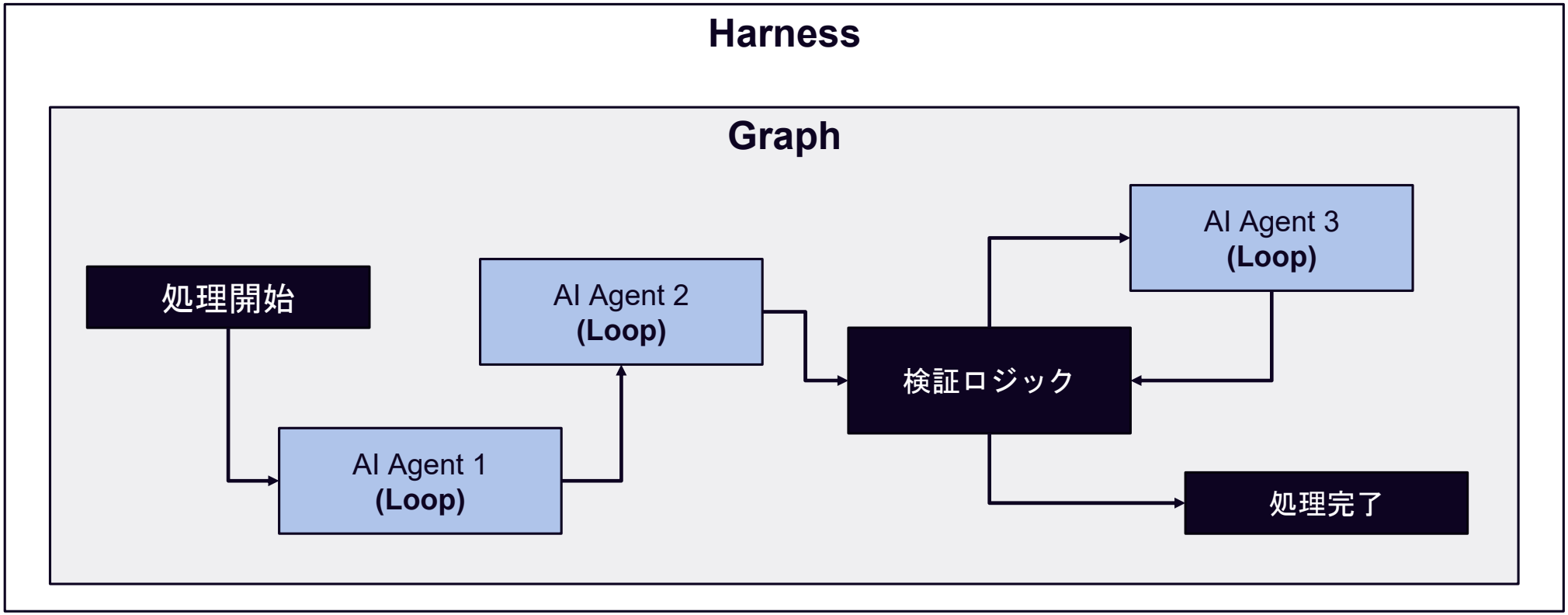
AI活用エンジニアリングの概念構成

- 基本的には、AIが動作するハーネスのなかにグラフ構造を作り、そのなかで自律的にループを回すことのできるAIエージェントを動作させてゆく形が基本となる。



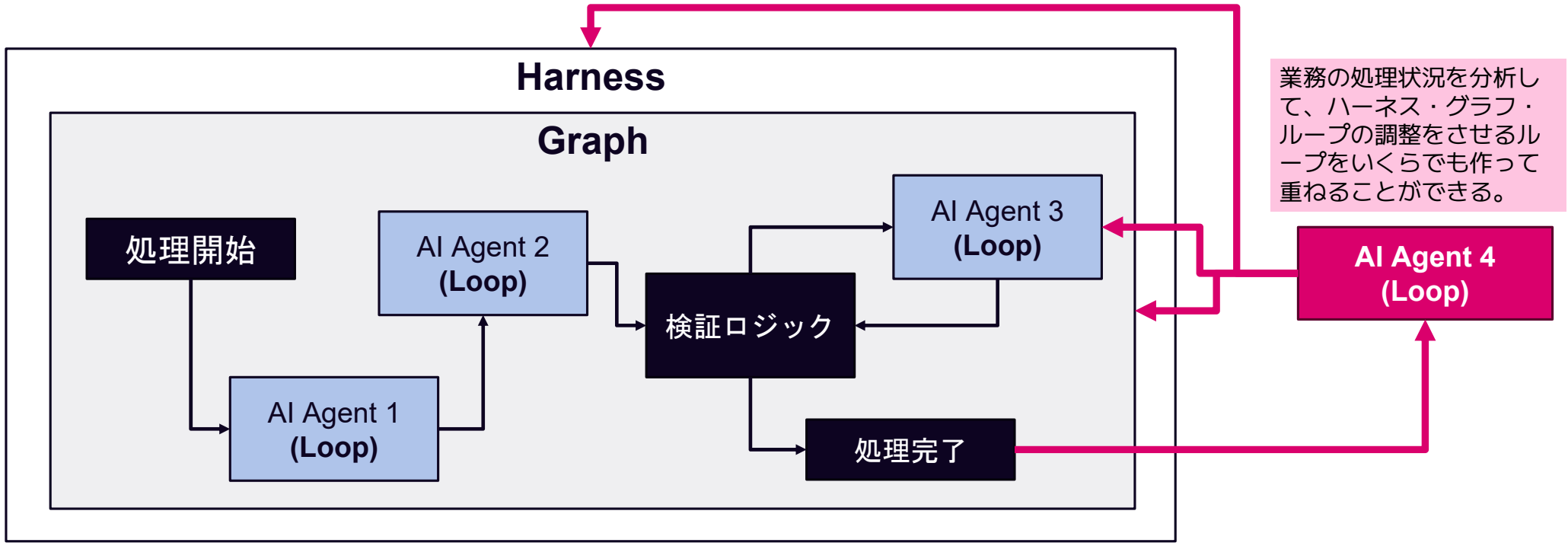
AI活用エンジニアリングの概念構成

- 基本的には、AIが動作するハーネスのなかにグラフ構造を作り、そのなかで自律的にループを回すことのできるAIエージェントを動作させてゆく形が基本となる。



ループエンジニアリングがなぜ重要なのか

- 素朴に考えると、ループエンジニアリングは、グラフエンジニアリングのなかで所定のタスクを受け持つAIエージェントの作り込みだと考えがちである。
- しかし、ループエンジニアリングは**グラフやハーネス、ループの改善に使える**「メタ性」を持つ。



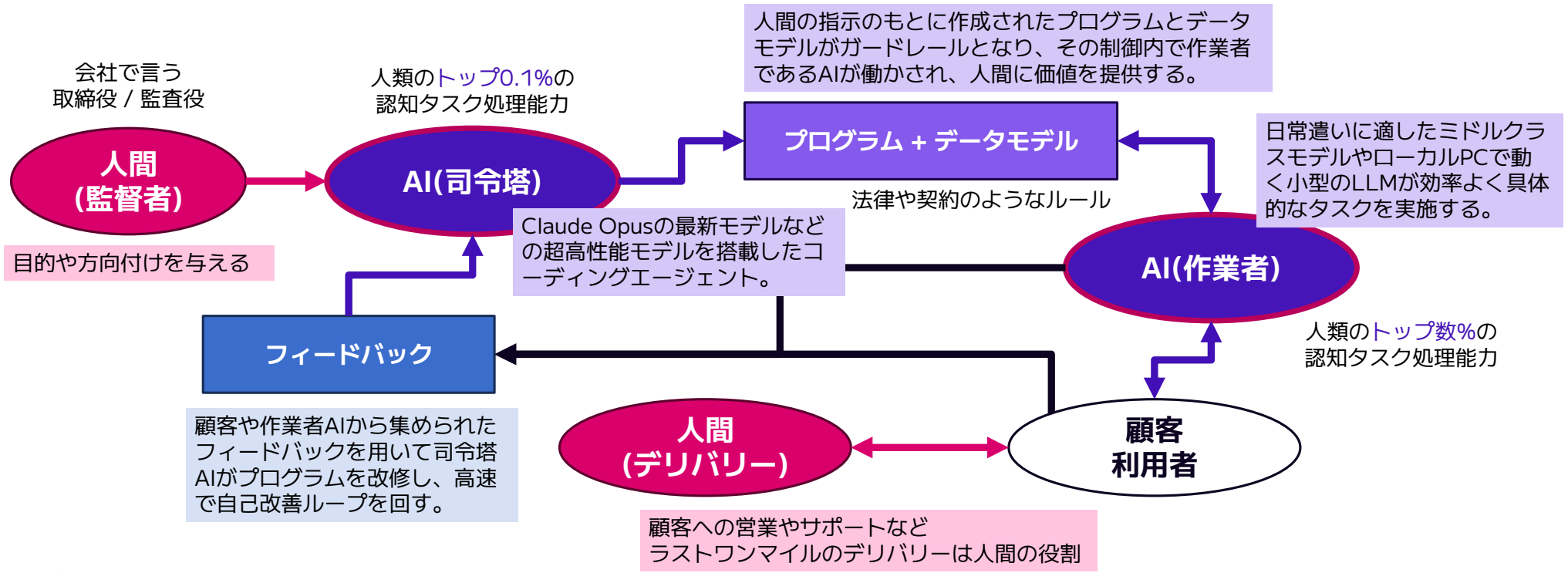
抽象度ごとのループエンジニアリングの活用法

- 特定のタスクを完遂させるための活用から、システムや事業の全体を視野に入れて改善を駆動させる方法まで、具体的な使い道にはかなりの抽象度の幅がある。

具体レベル	タスクレベル	• 与えられた日本語の文章を英語へ翻訳し、全体の文脈に沿って訳語が適切かを1語ずつ検討し、修正すべき点がなくなるまで改善を行い、最後に成果物の訳文と修正履歴のレポートを提出させる。 • 与えられた基準を満たすテストコードを書き、それらすべてのテストを通過するプログラムを書かせる。
	業務レベル	• 与えられた要件定義書に従って、必要なプログラムを漏れなく記述した上で、それらのプログラムが要件定義書にある要件をすべて満たすかを要件ごとに確認し、必要な修正を行い、全要件を満たすプログラムが完成したら、最終成果物と業務レポートを提出させる。
抽象レベル	アーキテクチャレベル	• AIチャットボットのロジックを改善するために、1000件のユーザーによる質問のサンプルデータを与え、それらに対する回答の品質を評価するevalを作り、evalの評価を改善するためのアイデアを出し、実装し、evalを回して結果を分析するループを回させる。
	事業・経営レベル	• ユーザーから寄せられたサービス改善要望やクレームの内容と、ソーシャルメディア上でのサービスの評判を継続的に収集・分析し、競合他社の動向を踏まえて事業成長につながり得る改善策を立案し、事業部にフィードバックレポートとして送った上で、施策実施後の変化もモニタリングして効果を判定させる。

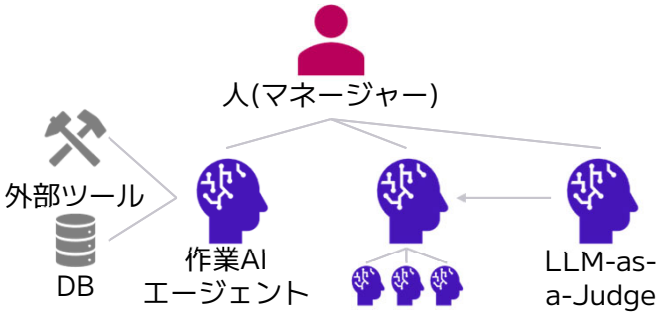
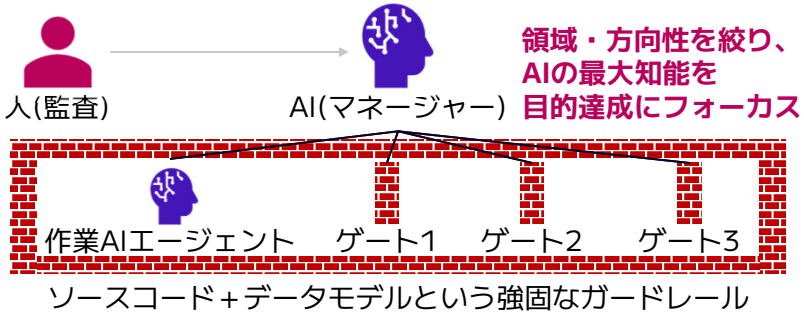
新コンセプト「フルレバAI全力2階建アーキテクチャ」

- AIの可能性を最大限引き出す仕事の仕方とは、「AIを使って、AIを使ってXXXする仕組みを作る」という手法で、私はこれを「フルレバAI全力2階建アーキテクチャ」と呼んでいる（”全力2階建”）。
- 今後は競争力を持つ組織・システムは**すべて**全力2階建アーキテクチャになると予想している。



人とAIの協働の変化イメージ

- 人がAIをマネジメントする段階から、AIがAIをマネジメントする仕組みを人がガバナンスを利かせて管理
- 高い知能を持ったAIを、平均回帰ではなく、期待値を満たす出力を再現性をもって発揮させるためには、
- 依頼者の固有性という偏りを強制的に出力させる必要があり、それがソースコードとデータモデル

	タスク支援レベル	自動化・半自動化レベル(確率論組織)	業務委任レベル(固有性を再現できる組織)
	チャットAI	人がAIエージェントをマネジメント	人がガバナンスを定め、 AIがマネジメントし、AIが実行
Input	• プロンプト • コンテクスト	• アーキテクチャ • プロンプトベースのルール・ガードレール • コンテクスト	• ガバナンスモデル • ソースコードベースのルール・ガードレール • データモデル • AIが人に求めた情報
Processing	 LLM (ブラックボックス化)		
Output	• 質問への回答 (ハルシネーション あり)	• 期待する品質に満たない**平均回帰した成果物** (再現性は乏しい) • プロンプトベースのため**ルール・ガードレール** **が守られないリスクあり**	• 期待する品質を満たす**「統計的に偏りがあり」かつ** **「再現性」のある成果物** • ソースコード・AIによる多段階チェックにより、 ルール・ガードレールを遵守

Graffer

【事例】基幹システム刷新Prjにおけるリバースエンジニアリング成果

- リバースエンジニアリングの多くの工程でAIが自動的・自律的に作業する仕組みを構築したことで、工程全体で、AIを用いない場合と比較して、93%の工数削減を実現

AI自動

プロセス改善

レビューの指摘事項を収集し、解析・出力プロセスを改善

-94%

0.5人日 → **15分**

AI自動

画面実装仕様書作成

- **解析**：AIが作成した調査ガイドを基に既存ソースコードを自動解析
- **出力**：規定のテンプレートに解析結果を構造化して出力

-98%

3人日 → **30分**

AI自動

AIレビュー

- **レビュー**：画面実装仕様書・ソースコード・Excel仕様書の3点をクロスチェック
- **自動修正**：AIが不整合検知、ソースコードを正として自動修正

-94%

1人日 → **30分**

AI自動+人間

当社レビュー

- **品質管理**：人間がExcel仕様書・実績画面を突合し、誤りや曖昧性・体裁を精査
- **調査・反映**：人間の指摘をAIが理解し、仕様書全体へ正確に展開

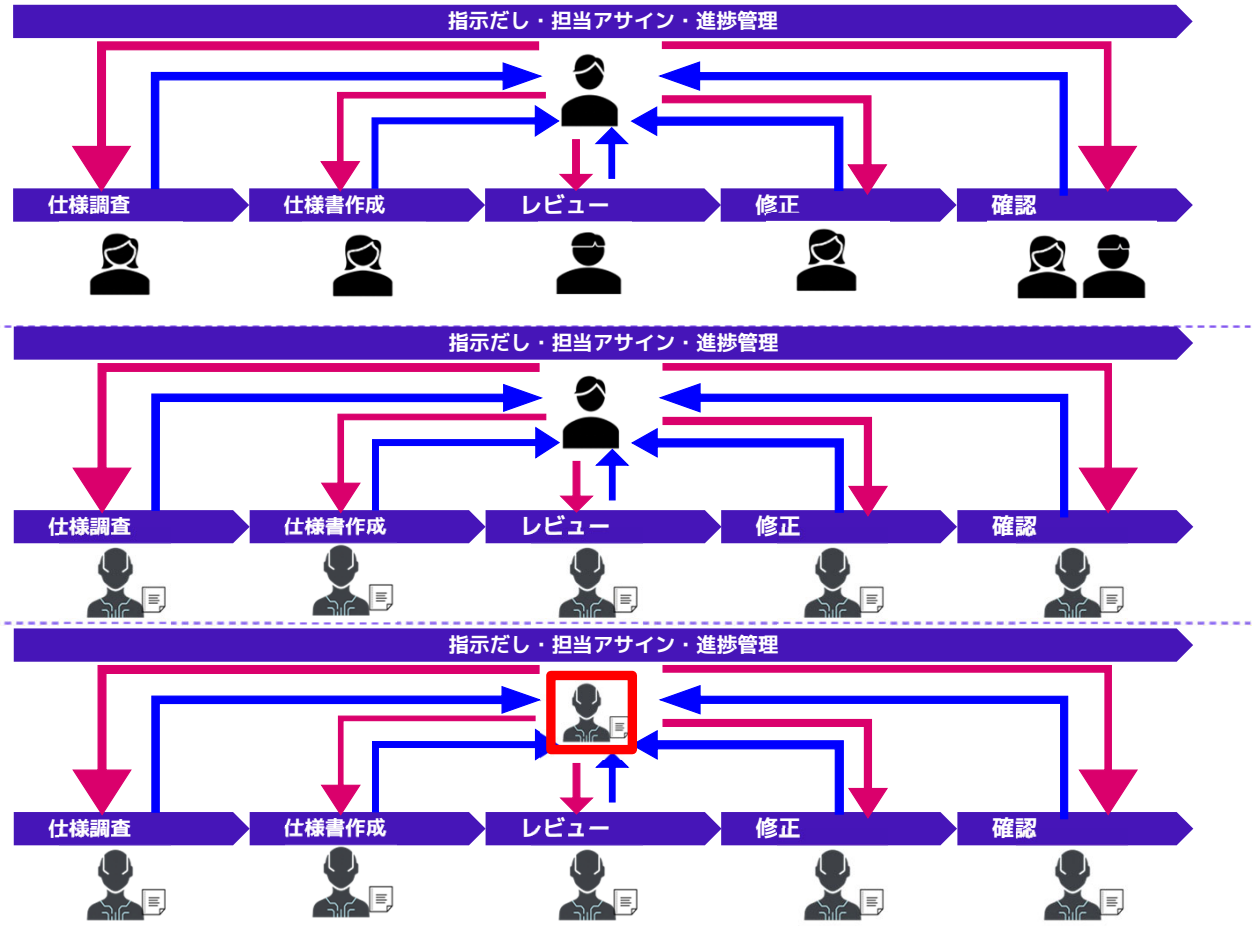
-81%

1人日 → **90分**

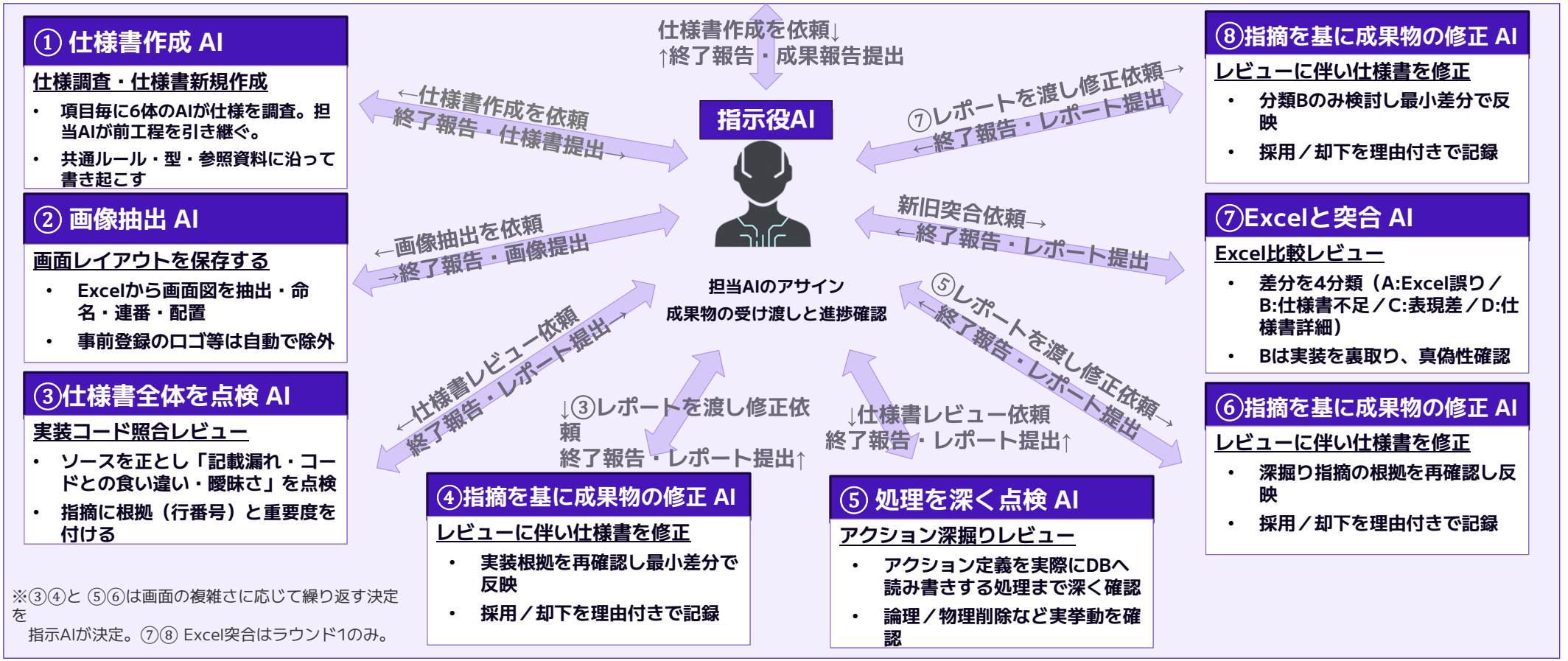
顧客レビュー

人がAIを起動する時代から、AIがAIを起動する時代へ

指示だし　完了報告

活用AIツール	変遷	担い手の変化
プロンプト • AIにその都度入力する具体的な指示文	phase1 人が人を動かす	
Skills • 調査手順・判断基準・出力形式を明文化した作業マニュアル群	phase2 人がAIを動かす	
AIエージェント • 対応するSkillを読み、目標達成のために自律的に作業を進める存在 **例：作業者・レビュワー**	phase3 AIがAIを動かす	
オーケストレーションAIエージェント • 起動順序・作業結果の受け渡し、各AIの進行状態を管理する進行役AI **例：プロジェクトマネージャー**		

プロジェクトのパイプライン全体図（例）



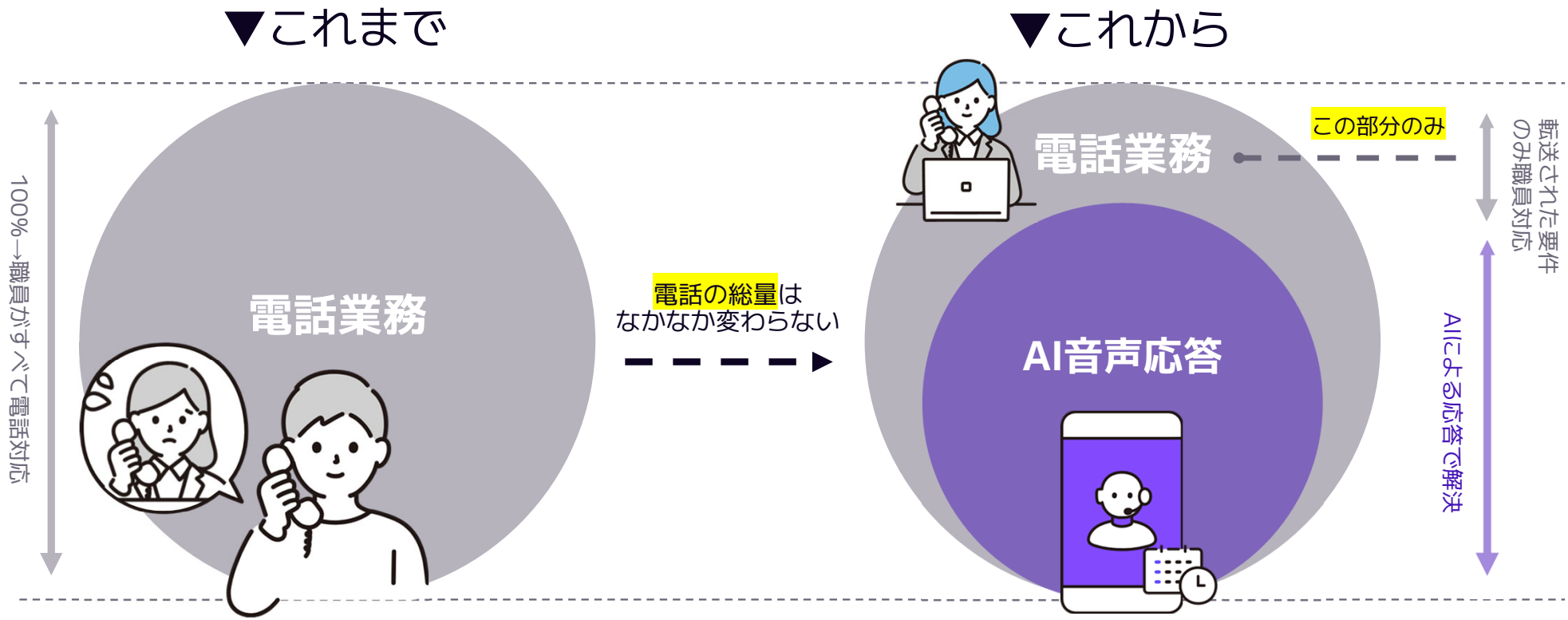
※③④と ⑤⑥は画面の複雑さに応じて繰り返す決定を
指示AIが決定。⑦⑧ Excel突合はラウンド1のみ。

Graffer レビューを繰り返しながら完走し、最後に人間の要確認事項をサマリーに集約して出力

行政デジタル化の事例

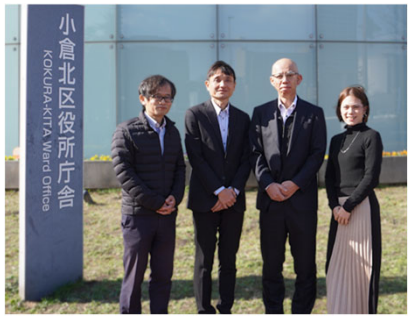
AIによる電話対応サービスが実用可能に

AIのマルチモーダル性能が向上した結果、電話応対についても一部をAIに代替可能。
既に自治体でも複数の活用事例が出てきており、「AI電話」は実用フェーズに入っている。



【事例】福岡県北九州市|マイナンバー関連の電話の約半数を削減

FAQをAIオペレーターにインプットした結果、マイナ関連の電話の約半数がAIオペレーターで完結*。
閉庁時間帯もAIオペレーターが質問回答を行うことで、市民利便性の向上にも寄与しています。



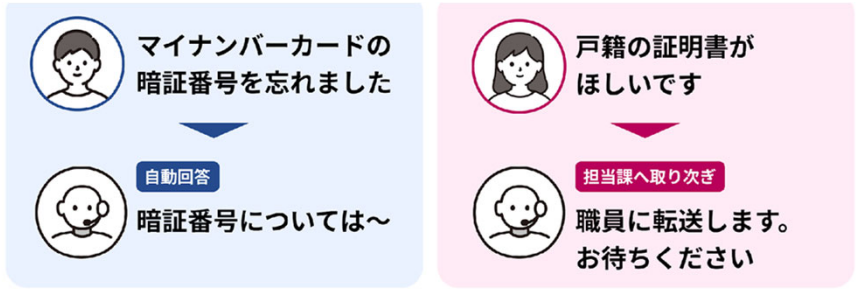
自治体	福岡県北九州市
導入課	小倉北区役所市民課
人口規模	90万人

▼導入前の課題

- 市民からの電話がなかなかつながらない
- 電話応対により職員の業務が中断される

ハルシネーション対策※

AIが回答できる範囲を明確に線引きすることで回答精度を担保



回答できる範囲とできない範囲について、事前に明確に線引きを行いました。例えば、「マイナンバーカードの代理受け取り」や「特急発行」といった、個別の判断が必要なケースはAIが回答せず、最初から職員に取り次ぐ仕組みにしています。

導入前に最も懸念していたハルシネーションについては現在のところ指摘は入っておらず、正確な案内ができているという点では、ひとまず安心しています。

※ハルシネーション：AIがもっともらしい嘘の情報を伝えてしまう現象

AIと職員が「補完し合う」未来へ



AIの立ち位置は「一次対応」であり、その先の「個別対応」は職員が担う。この役割分担をさらに明確にしていきたいと考えています。

自動化によって生まれた時間を、福祉分野や複雑な相談対応など「人が人に向き合うべき業務」にあてていくための仕組み作りが不可欠だと考えています。

*2025年9月の受電データに基づく

（参照）：https://graffer.jp/govtech/articles/kitakyushu-ai-call

【事例】愛知県稲沢市｜ごみの持ち込み等のお問い合わせにAIが回答

環境センターにおける受付時間、料金、持ち込めるごみの種類といった定型的な問い合わせに対して
AIオペレーターが回答。約9割の電話がAIで完結*しています。



*2025年11月～2026年2月の受電データに基づく

▼実証実験開始の背景
稲沢市ではデジタル化を推進する一方、ごみの持ち込みなど環境施設課への電話応対に時間を取られると、職員個々の本来業務に集中できないという問題や、異なる職員が対応するため電話応対の品質にばらつきが生じ対応品質の均一化が難しい状況が課題でした。このたび『Graffer AI オペレーター』の実証実験により、AIが定型的な質問に対応することで、市民の利便性を高めるとともに、職員はより丁寧な相談対応に集中できる体制の実現を期待しています。

（参照）当社ニュースリリース： https://graffer.jp/news/6956

【事例】東京都港区｜区民課宛の様々な問い合わせをAIが回答

「令和７年度みなと新技術チャレンジ提案制度」において当社提案が採択され港区と協定書を締結。
区民課職員の電話負担軽減に向けて実証実験を実施しています。



AIによる電話自動応答・職員取次の実証イメージ

「Graffer AIオペレーター」が自然な会話で要件を聞き取り。AIによる回答と、会話内容を引き継いだ上での職員への転送を自動で振り分けます。



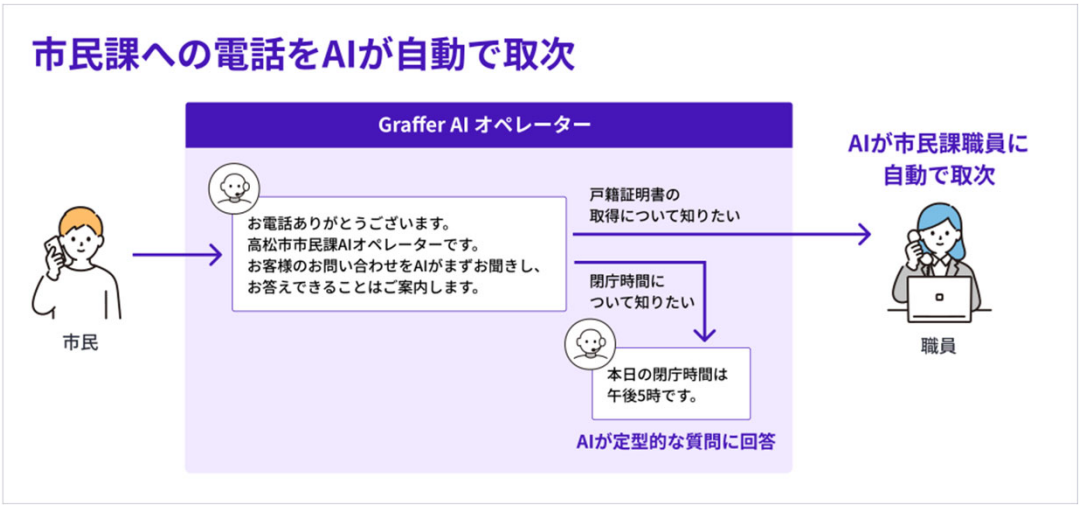
▼実証実験開始の背景
これまで港区では、ホームページやLINE公式アカウントでのAIチャットボットなど、デジタルでの案内を拡充してきました。しかし、電話は依然として多くの区民にとって最も身近な相談手段であり、転入・転出が集中する時期には区民課への電話が急増し、つながりにくさや職員負荷の増大が課題となっていました。市民サービスに関する問い合わせには、必要書類や手続きの流れ、窓口時間といった定型的な内容が一定割合を占めます。非定型で判断を要する案件は職員対応が不可欠である一方、定型問い合わせはAIエージェントが一次受けすることで、待ち時間の短縮や応答品質の平準化などが期待できます。

（参照）当社ニュースリリース： https://graffer.jp/news/6822

【事例】香川県高松市｜電話取次をAIで完結

市民課における電話取次をAIオペレーターが実施。電話の用件に応じた各係への電話取次をAIが実施することで、電話取次にかかる職員対応の軽減を目指しています。（2026年2月実証実験実施）



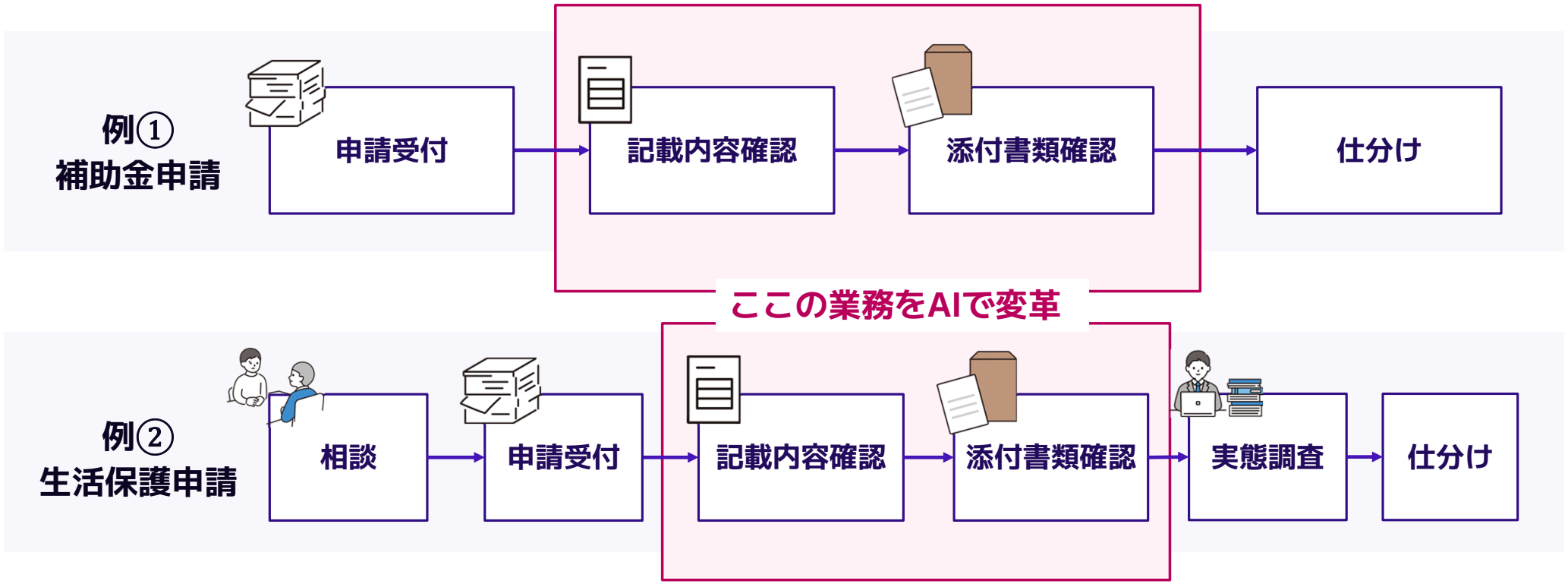


▼利用開始の背景
これまでも高松市では、行政手続きのオンライン化に取り組んできましたが、「市役所に電話してもなかなかつながらない」「夜や土日も問い合わせたい」という声が寄せられているのが現状です。特に市民課は、多いときで1分に数回程度の頻度で電話対応が発生することもあり、内容に応じて各係へ取り次ぐ運用が続いています。
このような背景から、高松市とグラファーは、AIエージェントを活用した電話応答サービス「Graffer AI オペレーター」を導入し、問い合わせ対応の高度化と持続可能な行政サービスの実現に向けた実証事業を開始することといたしました。

（参照）当社ニュースリリース：https://graffer.jp/news/7075

オープンウェイトモデルの性能向上で審査業務でのAI活用が可能に

担当者が目視でチェックしている個所を、AIが自動で確認します。



AIによる書類の審査自動化のデモ

デモ動画

審査ルールをもとに自動審査

設定したルールに基づき、プログラムが自動で合否を判定します。

不備ありの項目は判別可能

不備がある場合は色を変えて、視覚的に判断しやすいようになっています。
審査結果に基づき、正常・異常のステータスを自動付与。異常系レコードは一覧上で即座に判別可能

目視確認サポート

OCRで読み取ったデータと帳票データを比較することができます。もしOCRで読み取ったデータに誤りがあった場合は修正が可能です。

最終的に人がチェックしやすく、OCRや審査に誤りがあってもリカバリーできる設計

AIによる書類の審査支援エージェントのデモ

- 1:14
  書類の読み込み
  OCR
  審査

デモ動画

ローカル完結型の提出書類からの自動情報抽出ワークフロー

- 見つけたい項目をあらかじめ与えた上で、不定形な書類の写真をAIに渡す。
- AIは写真を見ながら、どこに何が書いてあるかを自律的に把握し、目的となる情報を探し、確信度が一定以上に高まるまで検討を続ける。
- 確信度に応じて結果を出力・報告する。

デモ動画

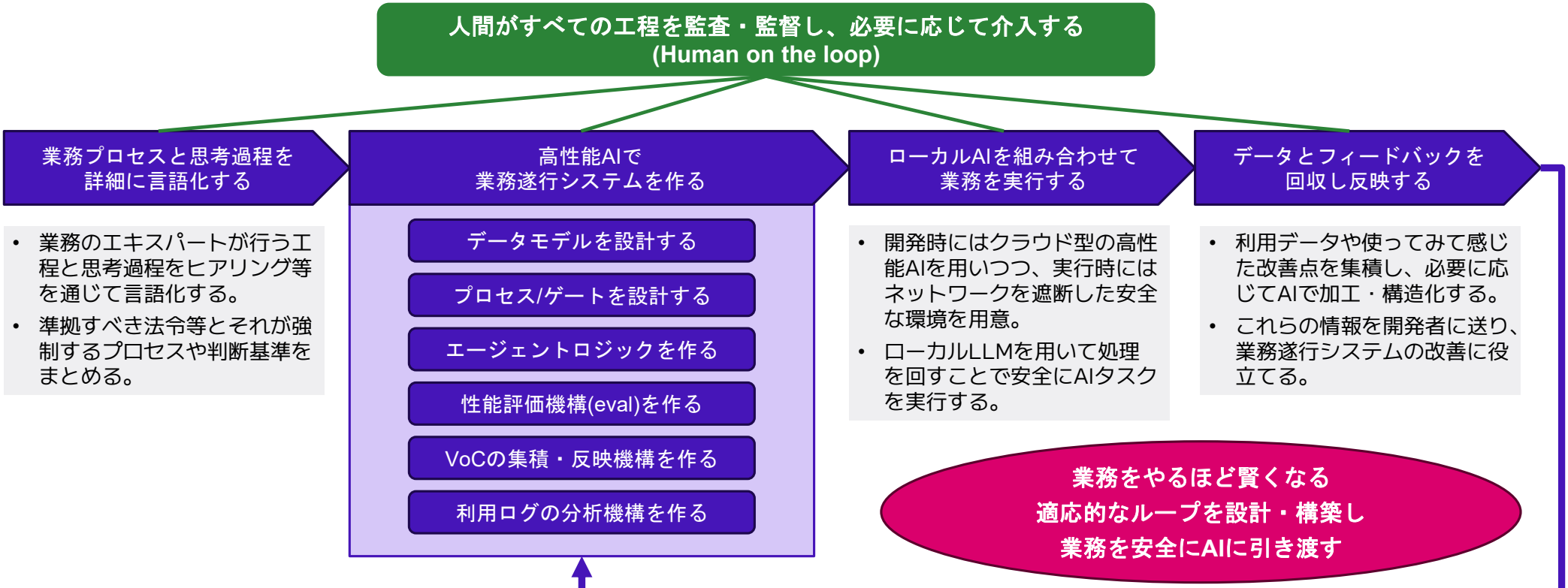
ローカル完結型の個人情報自動保護ワークフロー

- ローカルで動くLLMとOCRアプリを組み合わせ、個人情報が記載されている個所を画像から自動認識し、黒塗りした画像を作成する。
- これにより、個人情報記載のない画像を作成し、後続のワークフローで安全に活用することができる。

デモ動画

行政運営そのものを “ソフトウェア” で回す時代の到来

- 「AIを使って、『AIを使って●●する仕組み』を作る」という全力2階建のコンセプトを導入することで、ほぼあらゆる種類の認知タスクをAIを主体として実行しきることが可能になる。



まとめ

- AIによるコーディング能力が突出して進化することによって、コーディング以外の様々な業務もAIが設計・管理できる状況が到来している。
- 具体的には、「AIを使って、AIを使って業務を回す仕組みを作る」というAIを2段組するアーキテクチャを用意することで、人間の適切な監督下でほぼあらゆるデスクワークをAIによって遂行することができると考えられる。
- ソフトウェア開発の領域では、既に”Software Factory”といった概念が普及し、AIによる自動開発工程が実用化され、生産性向上が確認されている。
- 行政の領域でも、電話応対の自動化や書類審査の自動化については自治体において取り組みの事例が急速に増えている。