AIエージェントの定義と構成要素

Chapter 1 Section 2 | AIエージェント・ストラテジスト試験対策

今日学ぶこと

1. **生成AIとAIエージェントの違い**
   1往復の道具 vs 自走する存在
2. **三位一体モデル**
   頭脳・記憶・手足の3つで動く
3. **自律性の4段階**
   反応型 / 支援型 / 協働型 / 自律型
4. **支援型 vs 協働型の判別キー**
   試験で繰り返し問われる論点
5. **「何を任せないか」の判断**
   不可逆操作はHITL必須

道具として使うAIと

自走するAI。

何が違うのか？

ツール vs エージェント

ツールとして使うAI

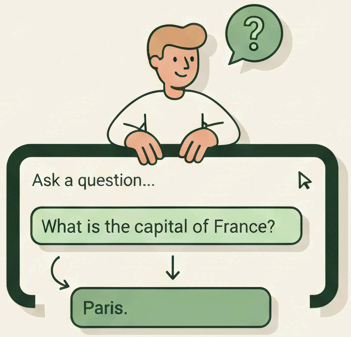
1往復のやり取りで終わる。次に何をするかは人間が考える

例：領収書を1枚だけ「経費カテゴリに分類して」

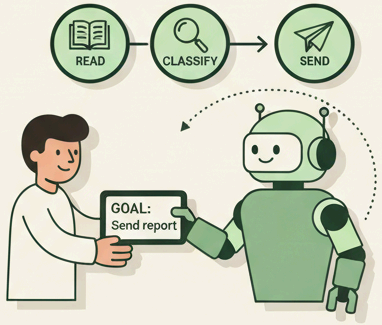
エージェントとしてのAI

目標を渡すと自走する。計画→ツール操作→実行→結果確認まで一気通貫

例：領収書画像を送ると「読み取り→分類→申請書作成→上長承認依頼」まで完結



TOOL: 1-TURN



AGENT: AUTONOMUS

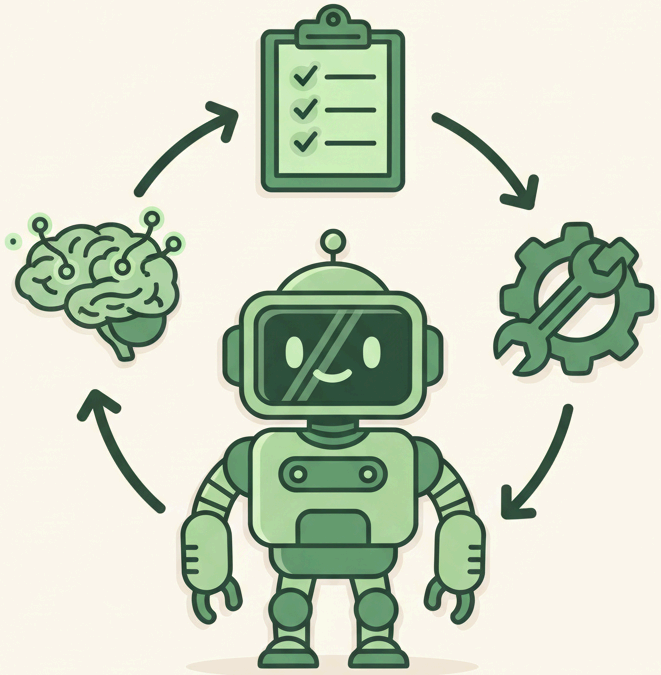
ツール（1往復）

エージェント（自走）

AIエージェントの定義

- 人間の指示（目標）を受けて**自律的に考え・判断し・行動する**AIシステム
- 計画立案・ツール使用・実行・修正までを**自分で繰り返す**
- 「聞けば答える道具」から「目的に沿って業務を進める支援役」へ

道具から、業務を進める存在へ



構成要素：三位一体モデル

頭脳
LLMが判断・推論

記憶
ナレッジベースを参照

手足
外部ツールを操作

3つのうち1つでも欠けると、AIエージェントとして**物足りなさが生じる。**

① 頭脳＝LLM

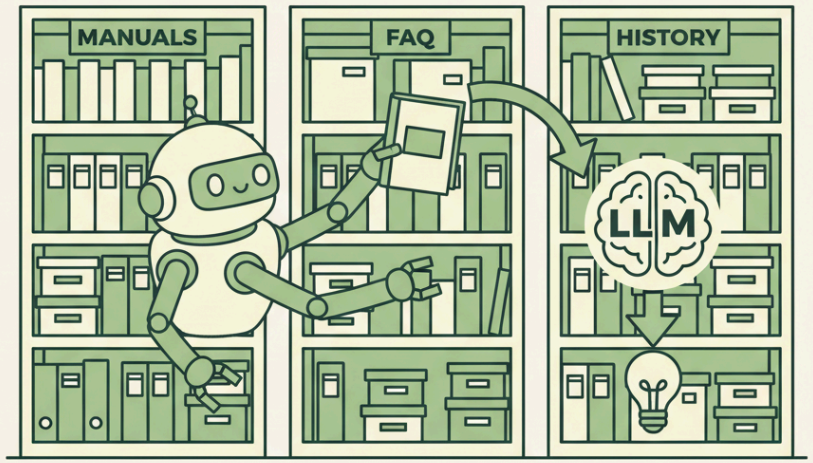
- 指示を理解し、**判断・推論**を行うエンジン
- 前回学んだ**LLMがここに該当**する
- どのLLMを選ぶかで、**コスト・精度・速度**が決まる

> 軽量モデルと高性能モデルを**タスクの難易度で使い分ける**のがモデルルーティング（Section 1）。

② 記憶＝ナレッジベース

- 業務マニュアル・FAQ・対応履歴など**社内情報**をAIに参照させる仕組み
- 記憶がないと、社内固有の質問に対し**汎用的な回答しか返せない**
- 参照の典型手法が**RAG**（後のSection 14で深掘り）

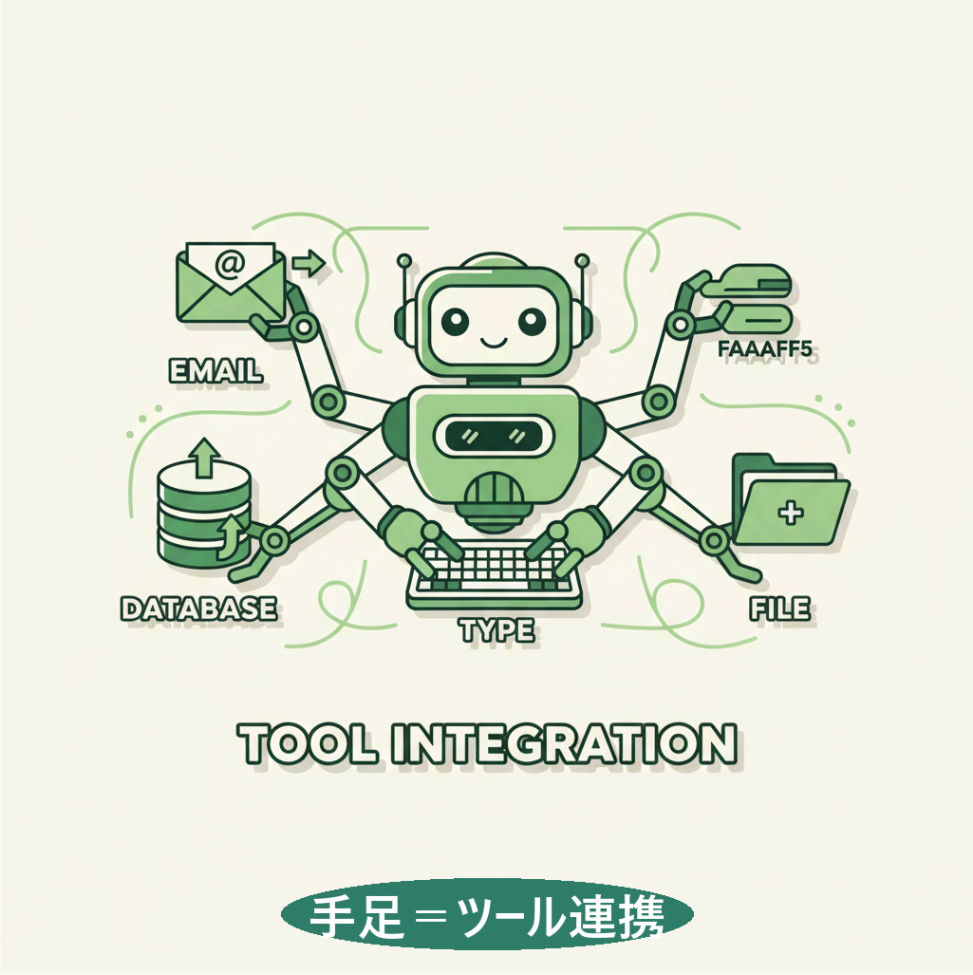
記憶＝社内データ。性能を決める源泉



③ 手足＝ツール連携

- メール送信・DB更新・ファイル生成など、**外部システムを操作**する機能
- 手足が無ければAIは**「考えて終わり」**
- 接続標準は**MCP**（後のSection 5で深掘り）

> 「申請書を作成してください」と頼んでも、手足が無ければ**「申請書の内容を提案しますね」**と返すだけ。実行できない。



1つ欠けるとどうなるか

欠ける要素	起こること
頭脳が欠ける	ルール通り動くだけの**自動化ツール**になる
記憶が欠ける	毎回ゼロからのやり取り、**社内固有の文脈ゼロ**
手足が欠ける	「高機能なFAQ検索」止まり、**実行できない**

自律性の4段階

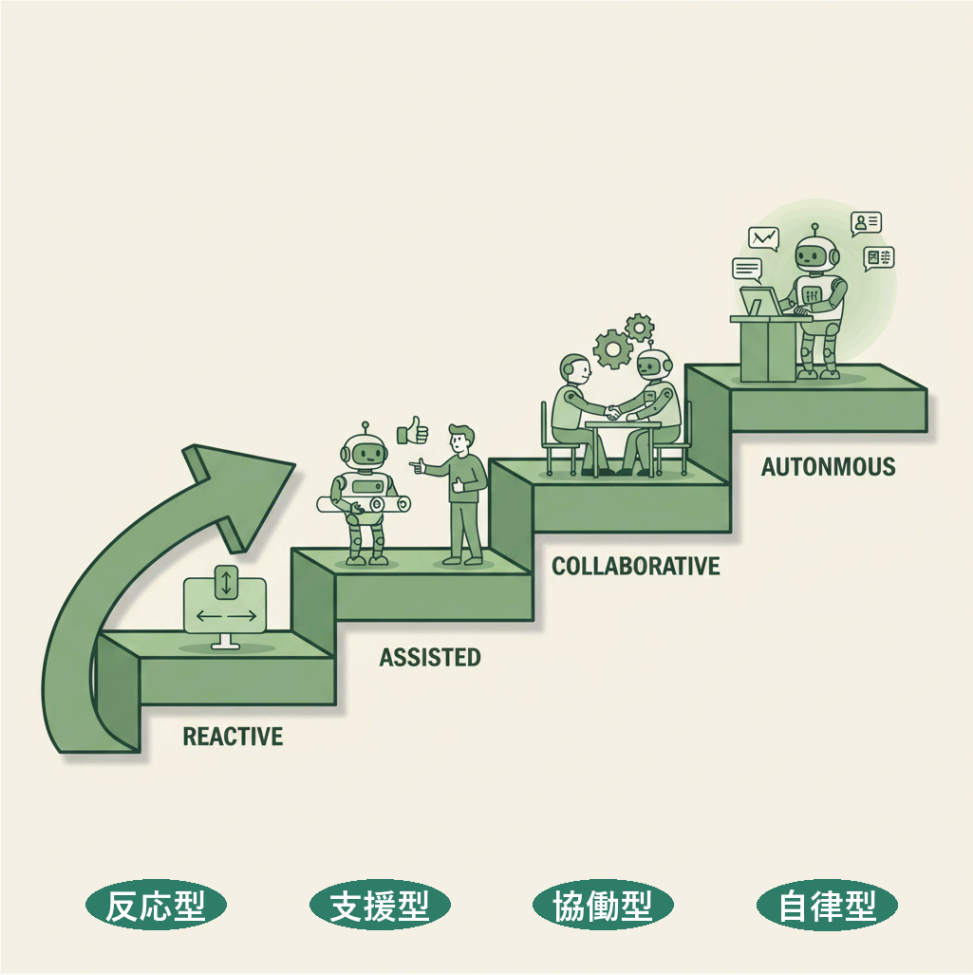
1
反応型
入力に即応するだけ。学ばない。Google翻訳など

2
支援型
案を提示。最終判断は必ず人

3
協働型
一部はAI自律判断、残りを人

4
自律型
目標から計画・実行まで完遂

人がどこまで関与するかで分かれる。**いきなり自律型を目指さず段階的に**引き上げるのが実務の鉄則。



レベル1：反応型

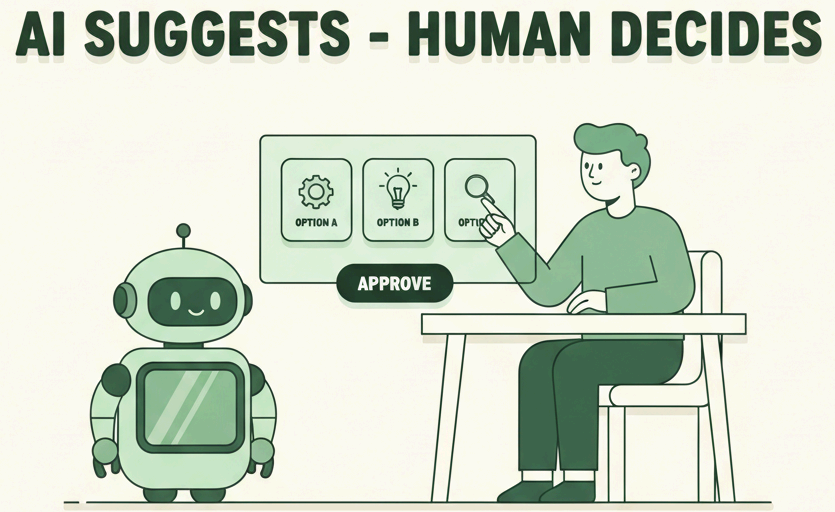
- 入力に対して**即座に出力を返す**だけ
- 過去から学習しない、自発的に動かない
- 例：Google翻訳、電卓のような決定論的応答

> エージェントというより「**反応する関数**」に近い。AIエージェント設計ではあまり登場しないが、対比の基準として押さえる。

レベル2：支援型

- AIが**回答案を提示。最終判断は必ず人間**
- 例：回答案を3パターン出して、担当者がどれを送るか選ぶ
- リスクが低く、**導入初期の標準**

人間が最後にYES/NOを押す設計

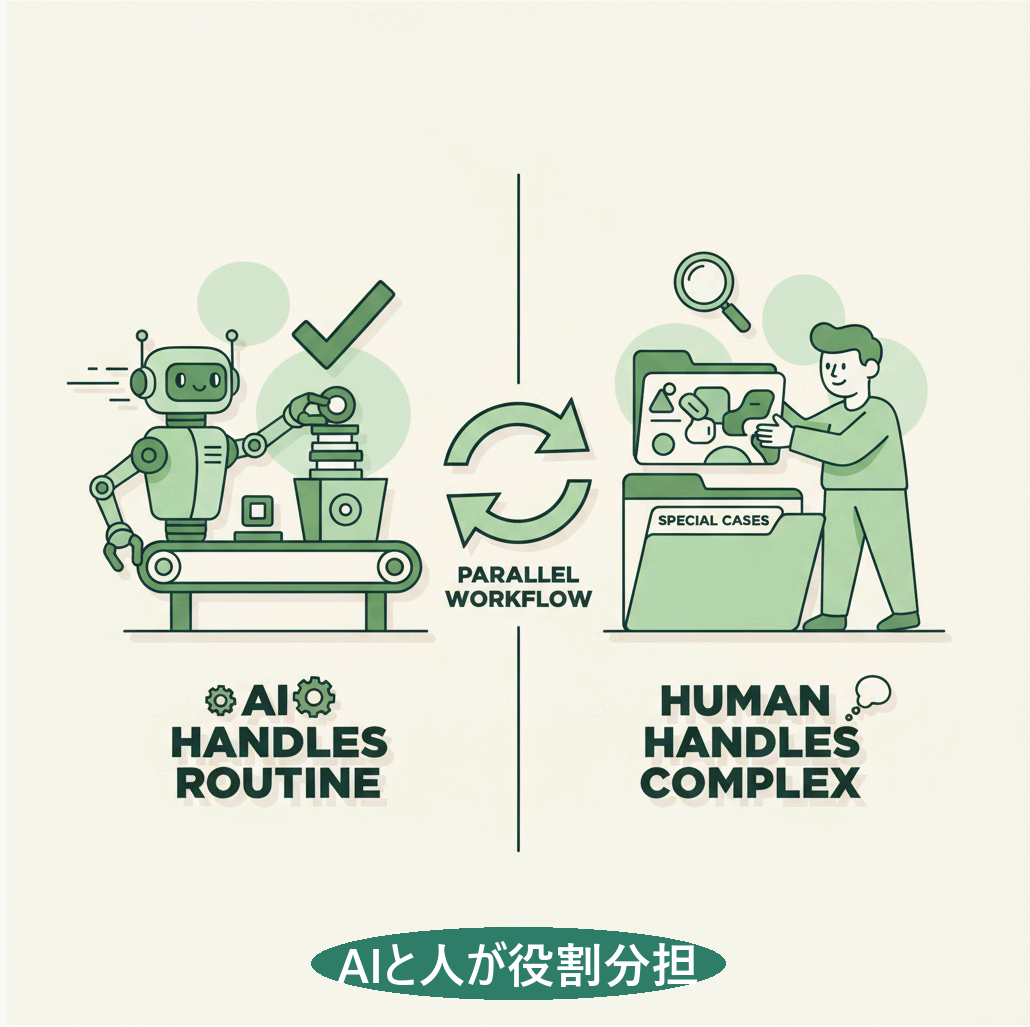


AIが提案・人が決定

レベル3：協働型

- 一部のタスクは**AIが判断から実行まで自律的に**完了
- 複雑な案件のみ**担当者にエスカレーション**
- 例：FAQ該当はAIが自動回答、それ以外は担当者へ

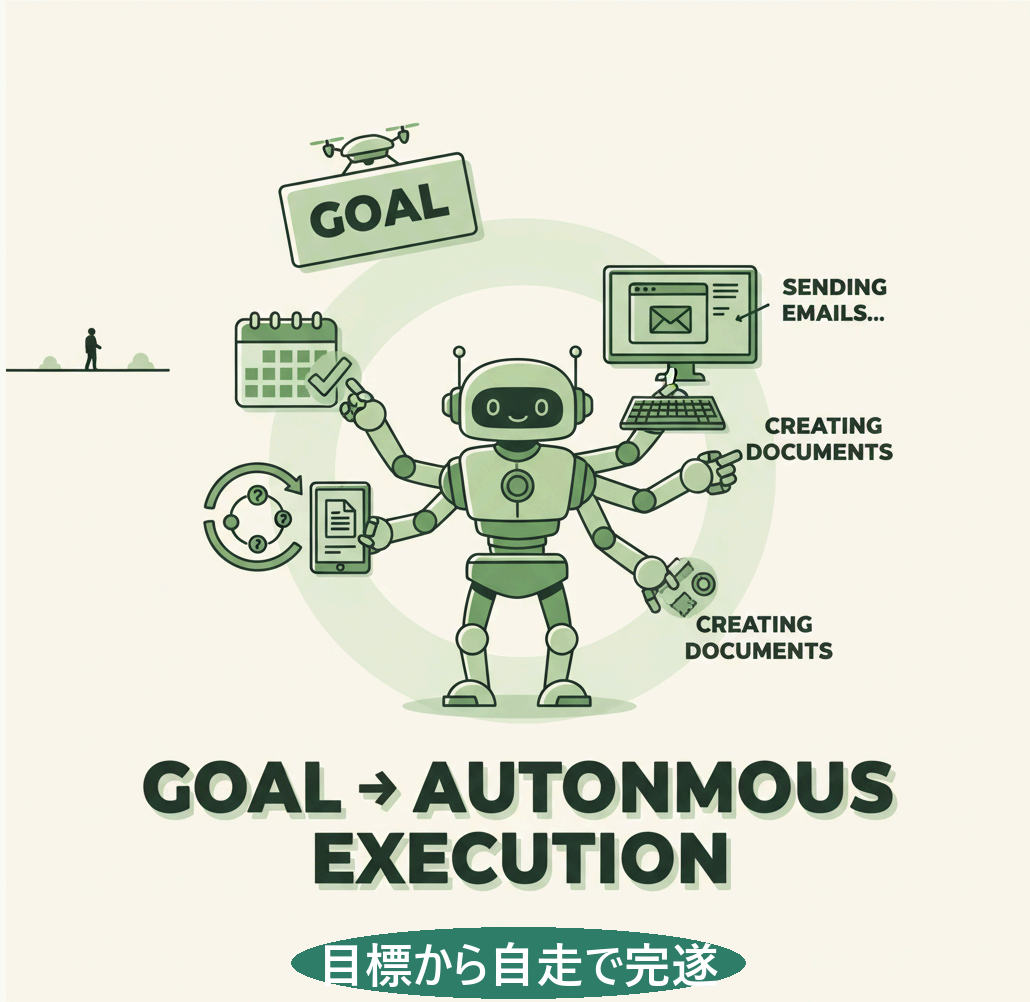
> 支援型から協働型への引き上げは「**定型部分の自動化承認**」を意味する。誤りリスクとのトレードオフ。



レベル4：自律型

- 目標を与えるだけで**計画立案→実行→修正**まで完遂
- 例：「来月のイベントの準備をして」で情報収集→調整→資料作成→連絡まで自動
- 技術的には実現可能だが**判断ミスの影響が大きい**

業務の重要度に応じて段階的に引き上げる



試験頻出：支援型 vs 協働型の判別

AIが「最終判断を含む意思決定」を自律的に行っているか？

- **支援型**：人間が必ず最終決定を行う（AIは下書きまで）
- **協働型**：AIが一部業務について自ら判断＋実行まで完了

> 例：すべての問い合わせに案を出して担当者が送信判断＝**支援型**。FAQ該当は自動送信＋複雑案件のみ担当者→**協働型**。

シナリオ：ヘルプデスクAIの現状業務

体制と数字

全社**3,000名**のIT部門ヘルプデスク
月**500件**の問い合わせを**3名**で対応
FAQ該当：月300件×15分＝**月75時間**
非FAQ：月200件×45分＝**月150時間**
合計**月225時間**消費中

日々の作業の流れ

メール受信→カテゴリ判断
FAQフォルダのExcelをキーワード検索
FAQ該当：手入力で返信／非FAQ：個別調査
対応記録もExcelに手入力
月次レポートは管理者が手作業集計

今ある課題：担当者の暗黙知に依存（属人化）／Excelキーワード検索だと該当検出が漏れる／対応記録漏れ月**20件**／月次集計に**8時間**

シナリオ：ヘルプデスクAIの設計判断

前提

現状業務は前スライドの通り
月**225時間**を**3名**で消費中

これから学ぶ3つの判断

三位一体で**設計要素を洗い出す**→自律性レベルを**どこに置くか**→何を**任せないか**

判断1
三位一体で
設計要素を洗い出す

判断2
自律性レベルを
どこに設定するか

判断3
「何を任せないか」
を決める

シナリオ：月500件のヘルプデスク

判断1：三位一体で設計要素を洗い出す

- **頭脳**：分類は軽量LLM・回答生成は高性能LLM（モデルルーティング）
- **記憶**：ITマニュアル・FAQ集・過去対応履歴をナレッジベース化
- **手足**：社員DB照会・Slack通知・チケット起票の連携

> 3つすべてを揃えて初めてエージェントとして動く。**手足の設計を省略すると、コピペが残り「FAQ検索ツール」止まり**になる。

シナリオ：月500件のヘルプデスク

判断2：自律性レベルをどこに置く？

- 月500件のうち**FAQ該当は約60%**＝定型・低リスク
- 残り**40%は非定型**＝個別判断が必要・誤り影響が大きめ

段階1

まず**支援型**で全件確認運用→精度の実データを取る

段階2

FAQ該当の精度95%以上を確認できたら**FAQだけ協働型**に引き上げ

シナリオ：月500件のヘルプデスク

判断3：「何を任せないか」を決める

判断基準は「やり直しがきくかどうか」

- **「会議室の予約方法を教えて」**→ AIが少しズレてもダメージ小→任せる
- **「全社通知を発信して」**→ 取り消し困難→**人間が必ず確認**
- **「退職者のアカウントを削除して」**→ 誤って在籍者を消したら復旧不可→**HITL必須**

試験で問われる5つの論点

1. **生成AIとAIエージェントの違い**（1往復 vs 自走）
2. **三位一体モデル**と1つ欠けたときの症状
3. **自律性の4段階**（反応型/支援型/協働型/自律型）
4. **支援型 vs 協働型**の判別キー＝「最終判断を自律で行うか」
5. **「何を任せないか」**＝やり直しがきくかで判定