



生成AIの現在地： 創造から自律へ

生成AIパスポート試験 第3章 完全要約 (シラバス2026年試験対応)

本資料は、テキスト、画像、音響、動画の生成技術から、RAG（検索拡張生成）、そして自律型エージェントに至るまで、最新の技術動向と倫理的課題を体系的に解説する学習用スライドです。

テキスト生成AI：言葉を理解する基盤技術



定義

自然言語処理 (NLP) と機械学習を用いて、文章やテキストデータを自動生成する技術。

NLPの役割

コンピューターに人間の言葉（自然言語）を理解させる技術。

LLM（大規模言語モデル）

NLPの規模を巨大化し、高度な文脈理解を可能にしたモデル。

画像生成AI：GANから拡散モデルへの進化



データ収集・前処理

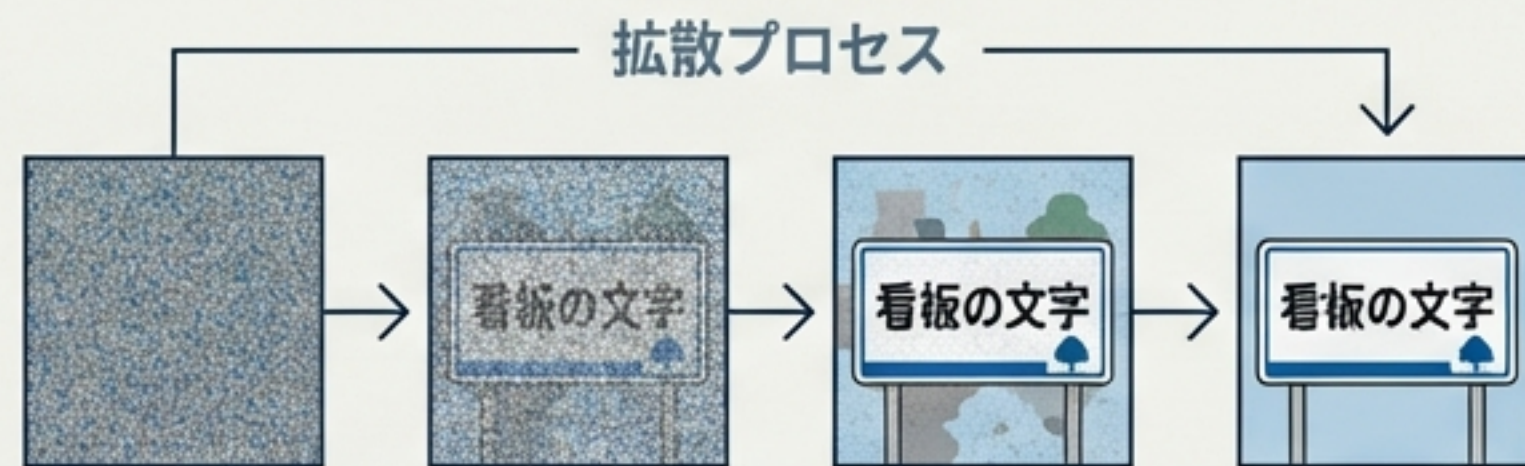
従来の手法 (VAE, CNN, GAN)

インターネット上の大量の画像データを収集・学習。前処理プロセスとして、画像のリサイズ、正規化、データの水増し (Augmentation) を経てモデルに適した形式へ変換。

モデル学習

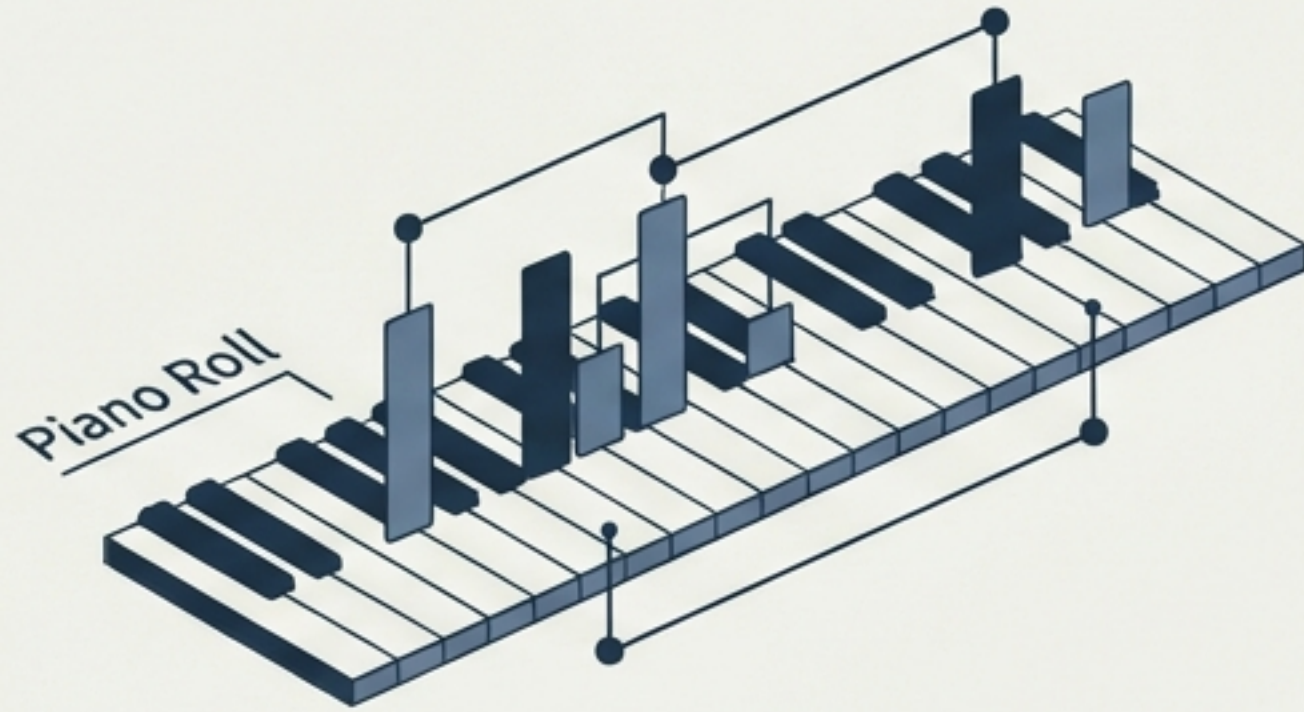
現在の主流 (Diffusion Models)

拡散モデルが中心。自己回帰型トランスフォーマー (Autoregressive Transformers) との組み合わせで高精細化。



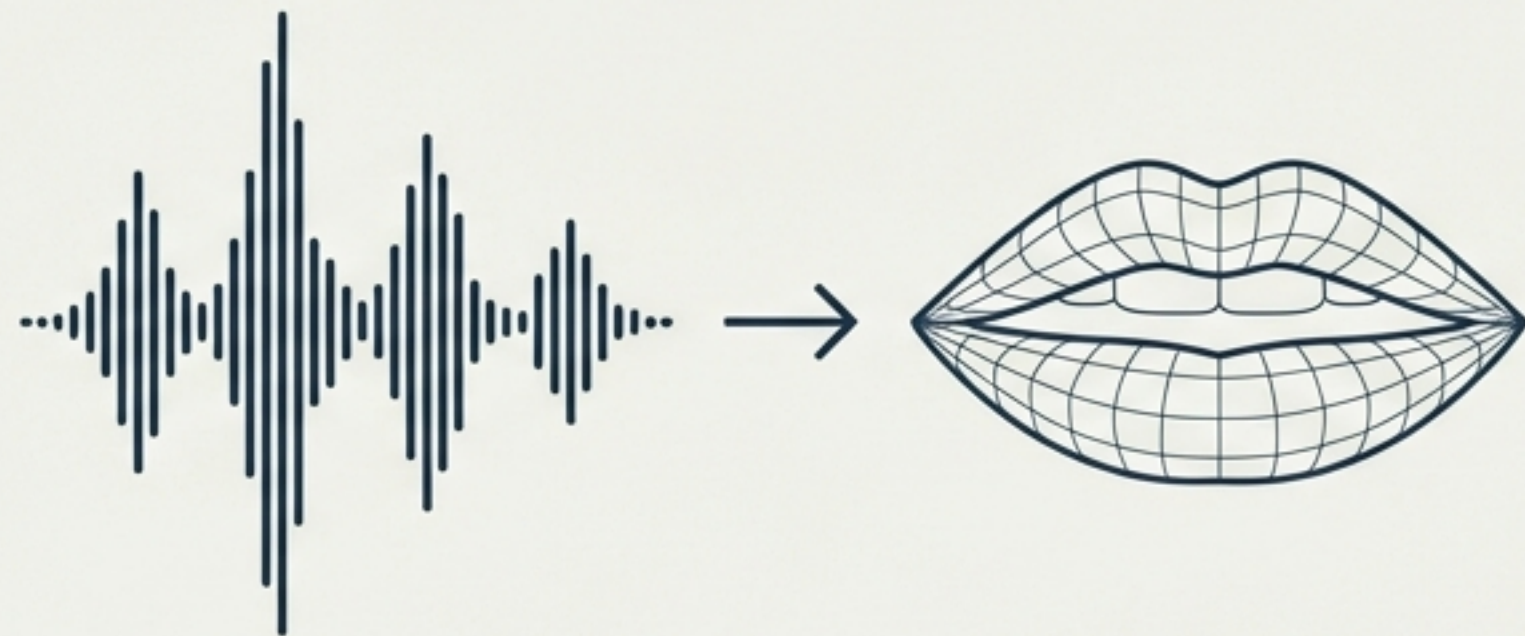
高精細化の鍵：画像をトークン化し順次生成することで、看板の文字や細線などの細部も破綻なく表現可能になった。

音声・音楽生成AI：聴覚データの再構築



音楽生成 (Music)

- 学習データ: MIDIファイルを含む既存の音楽データセット。
- 技術: シーケンス (連続) データの処理に長けたRNN (リカレントニューラルネットワーク) を用いて情報を水増し学習。
- プロセス: AI生成 → 人間によるフィードバック → 能力向上のループ。



音声生成 (Voice)

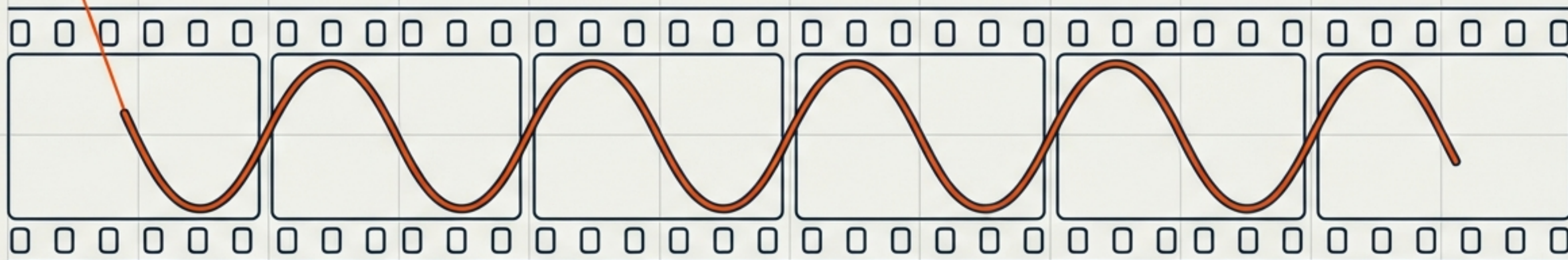
- 技術: ディープラーニング (教師あり学習) を使用。
- TTS (Text-to-Speech): テキストを自然な音声へ変換する技術。
- 注目ツール: ElevenLabs (少量のサンプルから声色を再現する高品質TTS)。
- Lip Sync: 特定の人物画像の口の動きを音声に合わせて動かす技術の進化。

動画生成AI：時間軸の整合性を保つ

仕組み: 静止画を連続的に結合して生成。

技術的課題: 各フレーム間の一貫性 (Consistency) を保つことが最重要であるため、RNNやVAEを利用して訓練される。

RNN / VAE Processing



Sora (OpenAI)

2024年2月発表。最長**1分間**の動画生成が可能。

Veo (Google DeepMind)

2025年5月発表。**超高精細映像**の生成と、**音声の同時合成**が可能。

生成AI導入のメリットとデメリット

メリット (Merits)

- ✓ - 作業効率の向上
- ✓ - 能力の補完と拡張
- ✓ - 新たなアイデア・デザインの創出
- ✓ - アクセシビリティの向上と、言語の壁の克服

デメリット (Demerits)

- ⚠ - 虚偽の情報（ハルシネーション）や不適切なコンテンツの生成
- ⚠ - 偏見・差別的なコンテンツの生成
- ⚠ - プライバシー侵害および著作権等の知的財産権侵害
- ⚠ - 特定業界の職業への影響と、質の低下

ディープフェイクと偽情報の拡散



定義:

ディープラーニングや生成AIを利用し、リアルな画像・コンテンツなどで人を「故意に騙す目的」で使われる技術。

社会的影響 (Social Impact):

1. ディスインフォメーション

フェイク映像が高品質化し、偽情報の拡散が社会的問題となっている。

2. ソーシャルエンジニアリング

詐欺や世論操作への悪用リスク。

RAG（検索拡張生成）：ハルシネーションの克服

従来のLLM



知識は学習データに限定

RAG



外部知識を検索・参照

概念 (Concept):

RAG (Retrieval-Augmented Generation): 外部の情報や関連文書を検索で集め、その根拠に基づいてLLMが回答を組み立てる技術。

重要性 (Why it Matters):

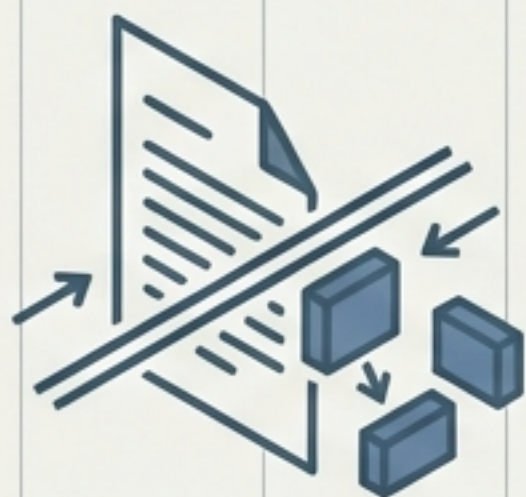
- ・ 従来AIの限界である「事前学習への依存」を脱却。
- ・ 最新の外部情報や関連文書を取り込むことが可能。
- ・ **ハルシネーション（もっともらしい嘘）の抑制に有効。**

ユースケース:

社内FAQ、規定検索、コールセンターでの手順案内（横断検索）。

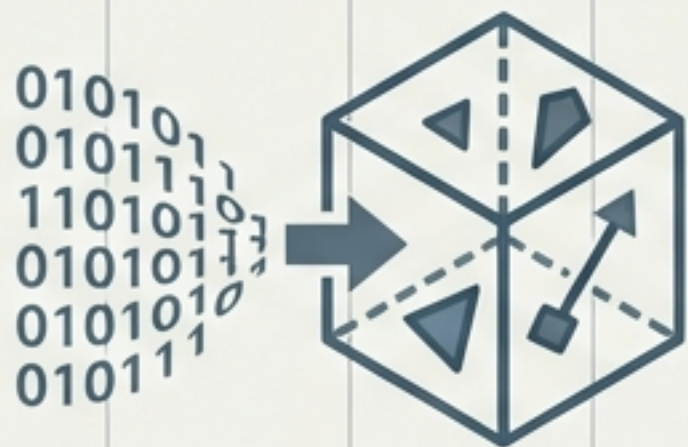
RAGの構築プロセスと仕組み

1. データ準備



チャンク分割：検索にかけやすいよう、文章を細かく分割する。

2. インデックス化



埋め込みベクトル：データの意味を数値化（ベクトル化）し、ベクトルデータベースに保存する。

3. 質問解析・回答生成



ユーザーの質問に最適なチャンク（情報の塊）を検索。LLMがその情報を参照して文章を組み立て、回答する。

AIエージェント：自律的なタスク実行

定義:

単なる回答ではなく、複数タスクを横断して処理するサイクルを持つ。

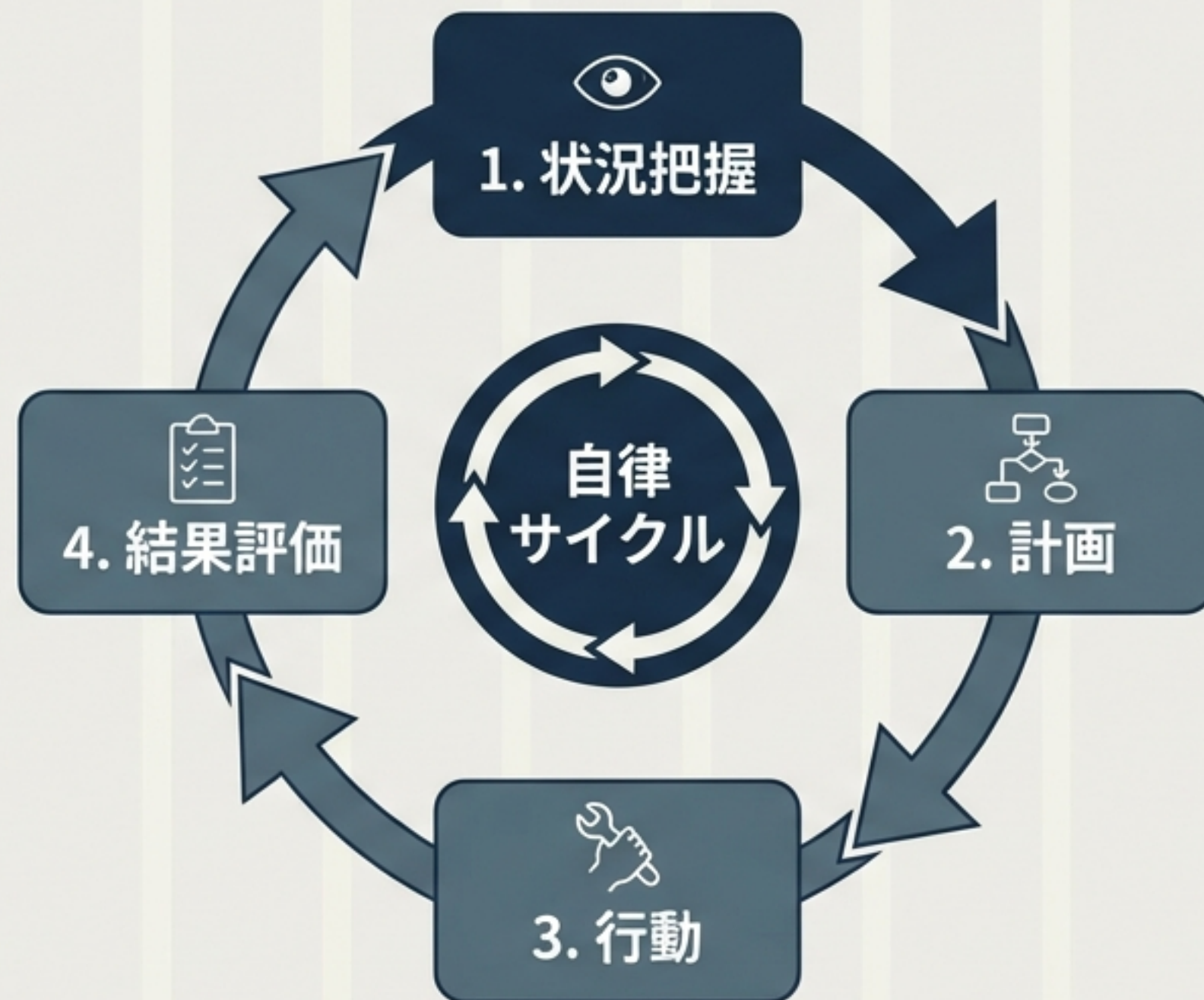
エージェントの種類:

1. ワークフロー型

受注処理や請求書発行など、あらかじめ定義した業務フローを処理する。

2. 自律型

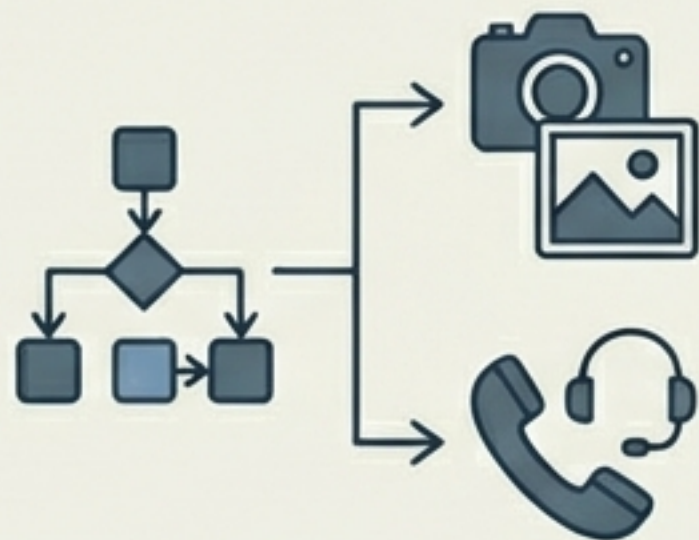
LLM・外部ツール・ループ構成を持ち、最終目標のためにタスクを「**再分割・再実行**」する。



AIエージェントの活用事例とツール

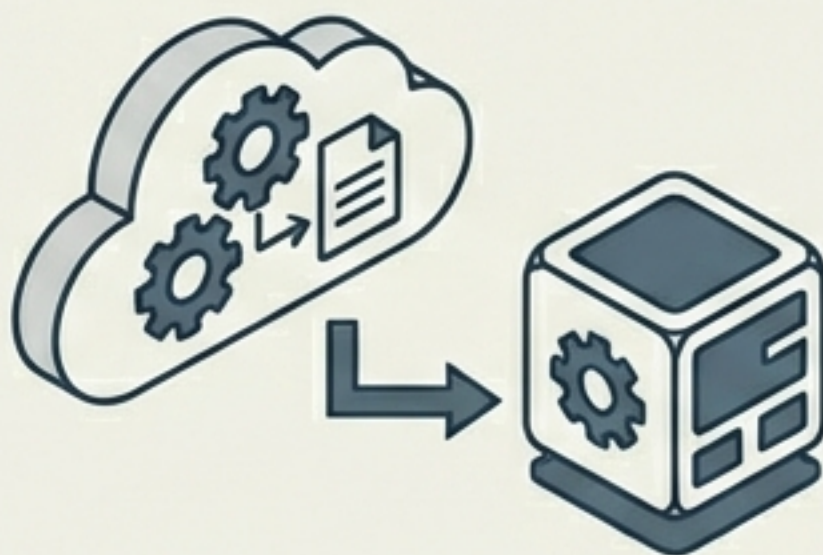
Genspark (All-in-One) (Helvetica Now Display)

- オールイン型スーパーエージェント。
- 工程管理、画像生成、電話代行など多様なタスクを実行可能。



Manus (General/Cloud) (Helvetica Now Display)

- シンガポール開発の汎用エージェント。
- クラウド上で独立稼働し、人間の操作なしで複雑な事務処理を全自動化する。



Skywork AI (Workspace) (Helvetica Now Display)

- ワークスペース型エージェント。
- 同じ作業空間で文書、スライド、表計算、音声などを複合的に作成する。



MCP：LLMと外部ツールをつなぐ標準規格



定義：

Model Context Protocol (Anthropic社発表)。LLMと外部ツール・データソースを安全かつ一貫した方法でつなぐ仕組み。

実行フロー：

1. 利用可能ツールとスキーマ（構造）の取得。
2. LLMによるツール選択。
3. コマンドの生成・送信。
4. 結果の受信・提示。

※これにより、自然言語の指示から外部APIを実行可能にする。

第3章 重要キーワードまとめ

Diffusion Model

現在主流の画像生成技術。ノイズ除去プロセス。

RNN

音楽・動画・音声など、連続データの処理に必須。

Deepfake

騙す意図を持った合成メディア。

RAG

外部データ検索によるハルシネーション抑制。

Embeddings

RAGで使用する「意味の数値化（ベクトル）」

MCP

Anthropic提唱のツール接続プロトコル。

AIエージェント

「状況把握・計画・行動・評価」のサイクルで自律稼働。