

はじめに

生成AIの使い方を論理的に説明できますか？

ChatGPTが登場して以来、世界ではずっと生成AIブームが続いています。生成AIの具体的な活用方法といえば、やはり文章の要約やアイデア出しなどが有名でしょう。実際、これらは生成AIの性質上、合理的な使い方といえます。

それでは、これらの使い方がなぜ有効なのか、あなたは説明できますか。あなたが本当に取り組むべきなのは、生成AIの表面的な使い方（How）を片っ端から調べて時間を浪費することではなく、**そもそもなぜその使い方が有効なのか、Whyを知ることです。**そしてこのWhyについては、**実は生成AIの登場から大きく変わっていません。**

基本に立ち返って理屈で考えることで、見かけ上のちがいに振り回されず、冷静に生成AIの活用方法を見定められるようになります。なお、ここであげた例については、本編の2.3で解説します。

忙しい人に生成AIの本質だけを伝える本が欲しい

筆者は業界でも珍しい「AI研究開発」と「システム開発」両方のリーダー経験を持つレアな人材として、生成AIの基礎知識から本格的なシステム導入まで1冊で体系的にカバーできるよう、**エッセンスだけを厳選しました。**

本書は、よくある「特定ツールだけに肩入れ・フォーカスする本」ではありません。フラットな立場から業界全体を俯瞰し、商用利用できる主要な生成AIツール群を幅広く紹介します。ただし、こうした特定ツールの使い方（How）は流行り廃りが激しく、実際このようなハウツー本はすでに数多く存在します。そのため、本書では各ツールの詳細なセットアップや操作方法までは扱いません。代わりに、**なぜそのツールを使うのか（Why）**に注力して紹介します。

同様に、本書では生成AIにまつわる基礎知識から応用技術まで一通り紹介しますが、各手法（How）を淡々と羅列するのではなく、理屈（Why）にもとづいて**どう使い分けるか（Which）**まで踏み込んで解説します。

●免責

本書に記載された内容は、情報の提供だけを目的としています。したがって、本書を用いた運用は、必ずお客様自身の責任と判断によって行ってください。これらの情報の運用の結果について、技術評論社および著者はいかなる責任も負いません。

本書に記載がない限り、2026年1月現在の情報ですので、ご利用時には変更されている場合があります。

また、ソフトウェアはバージョンアップされる場合があり、本書での説明とは機能内容や画面図などが異なってしまうこともあり得ます。本書ご購入の前に、必ずバージョン番号をご確認ください。

以上の注意事項をご承諾いただいた上で、本書をご利用願います。これらの注意事項をお読みいただくずにお問い合わせいただいても、技術評論社および著者は対処しかねます。あらかじめ、ご承知おきください。

●商標、登録商標について

本書に登場する製品名などは、一般に各社の登録商標または商標です。なお、本文中では™、®などのマークは省略しています。

誰もが生成AI導入を成功に導ける本が欲しい

本書では、明日から実践できる**具体的なアクションプラン**を提案します。生成AI活用のスタートダッシュを支援することはもちろん、筆者自身のAIシステム導入経験をもとに、「PoC止まり」で終わらないための注意点を具体的に紹介します。後者についてはITエンジニアだけでなく、業務ユーザーから経営者まで、特に**これから生成AI案件に関わるすべての人**に知ってほしい内容を詰め込んでいます。

また、本書は初心者でも概念をイメージしやすいよう、**図解にも力を入れています**。各章のまとめにも見開きの図解を入れているので、振り返りやイメージアップに活用してください。一部の図には**謎解き要素**や**小ネタ**が仕込まれています。ところどころに隠されたユーモアを探してみてください。

想定読者と本書のゴール

本書の主な想定読者を次に示します。

- 生成AI導入案件のポイントを手っ取り早く押さえたい方
- RAGやAIエージェントなどの生成AI技術を体系的に学びたい方
- ChatGPTの活用レベルを要約や文章生成から上げたい方

なかでも、本書は**生成AI導入案件の当事者**を強く意識しています。当事者というのはITエンジニアはもちろん、業務ユーザーや企画職も含まれます。筆者の経験上、こうしたステークホルダー全員が生成AIのポイントを押さえているプロジェクトは成功しやすいです。ただし、一部の開発者向けコンテンツに関しては、見出しに「上級者向け」とつけています。

あなたが本書を読み終える頃には、次の状態になっているはずです。

- 生成AIの基本知識を体系的に理解している
- ベースライン技術（RAGやAIエージェント）を自分で選定・説明できる
- 企業独自のデータやプロセスを扱う業務にも生成AIを活用できる
- 生成AI案件の特性に応じて、何を勉強すればよいか次の一手がわかる
(例：高度なRAG案件に参画する場合、検索エンジンや文書加工処理)

本書によって、日本全体の生成AI活用レベルがほんの少しでも底上げできるとしたら、AIエンジニアとしてこれ以上の幸せはありません。

謝辞

本書は、多くの方々のご支援なくしては完成し得ませんでした。日鉄ソリューションズ株式会社の園部勲さん、および山田一宏さんには、主に技術的な視点でレビューいただきました。また、同僚のみなさん、特に永井洵伎さんには、ITエンジニア目線で図解やストーリーについて助言いただきました。さらに、かつて長崎東高校という山の上で苦楽を共にした若村健人さん、小牧博也さん、西田知句さん、すばはむさんには、外資系コンサルタントから歯科医師まで多様な背景を持つ読者の視点から、構成や言い回し、楽しく読み進められる工夫について多くの示唆をいただきました。最後に、企画段階から何度も打ち合わせを経て、本書の完成にいたるまで支えていただいた中山みづきさん、ならびに技術評論社の方々に、心より感謝申し上げます。

目次

はじめに	iii
第 1 章	
まずは生成AIの基本を知ろう！	1
1.1 生成AIと大規模言語モデルって何？	2
1.1.1 ChatGPTの登場	2
COLUMN とはいえ、ChatGPTも万能ではない	3
1.1.2 生成AIの基本用語	3
1.1.3 言語モデルと大規模言語モデル	4
1.2 大規模言語モデルってどんな仕組みで動いているの？	5
1.2.1 Transformerの登場	6
1.2.2 大規模言語モデルの学習：事前学習	7
COLUMN GPTって何の略？	8
1.2.3 大規模言語モデルの学習：ファインチューニング	8
1.2.4 大規模言語モデルの学習：アライメント	10
1.3 大規模言語モデルは何が新しいの？	12
1.3.1 Scaling Law：サイズを増やすと精度が上がる	12
COLUMN LLMはMoney is All You Needの世界？	14
1.3.2 General-Purpose：複数のタスクを解ける	14
1.3.3 In-Context Learning：アドリブが効く	15

1.3.4 Chain-of-Thought：人の思考ステップを真似できる	18
1.3.5 Step by Step：自ら思考ステップを産み出せる	20
COLUMN 「think step by step」が不要な推論モデルの登場	22
1.4 まずは三大LLMサービスを覚えよう！	22
1.4.1 OpenAIのChatGPT	23
1.4.2 GoogleのGemini	23
1.4.3 AnthropicのClaude	24
1.5 生成AIを使うときの注意点は？	25
1.5.1 コンテキストウィンドウ：トークンと入力量の限界	25
1.5.2 ハルシネーション：もっともらしい嘘	27
1.5.3 オプトアウト：漏えいの予防策	28
1.5.4 不適切コンテンツ：暴力的・差別的な表現	29
1.5.5 プロンプトインジェクション：生成AIの攻撃	30
1.5.6 AI事業者ガイドライン：責任あるAIの共存 上級者向け	33
1.6 生成AIの著作権と向き合おう！	34
1.6.1 著作権侵害となる条件 上級者向け	35
1.6.2 AIの学習と推論のちがひ	37
1.6.3 AIの学習・開発における著作権 上級者向け	39
1.6.4 AIの推論・利用における著作権	40
COLUMN 生成AIの著作権侵害は誰の責任？	42
1.6.5 AIの生成物そのものの著作権	43
第 1 章のまとめ 生成AIの基本用語	46

第2章

すべてはプロンプトから始まる！ 49

2.1 プロンプトの基礎知識	50
2.1.1 プロンプトの基本要素	50
2.1.2 システムプロンプトとユーザープロンプト	54
2.1.3 温度パラメータ	55
2.2 プロンプトを書くときの基本ルール	57
2.2.1 否定文だけの指示は避ける	57
2.2.2 結論から書かせない	59
2.2.3 サンプルは直感的な表現で均等に与える	62
2.2.4 Markdown 記法で構造化する	63
2.2.5 Mermaid 記法でフローチャートを書く	66
2.2.6 XML 記法で複雑な情報を色分けする	68
2.2.7 LLMの学習データに形式をあわせる	72
2.3 生成AIの長所を活かして短所を補おう！	74
2.3.1 ケース1：文章を自然に整える	75
2.3.2 ケース2：裏取りの手間がかからない	76
2.3.3 ケース3：入力情報≧出力情報になる	78
第2章のまとめ プロンプトの基本要素	82

第3章

実践！
プロンプトエンジニアリング 85

3.1 たったこれだけ！プロンプト基本パターン10選	86
COLUMN コンテキストエンジニアリング	86
3.1.1 Zero-shot Prompting	86
3.1.2 Few-shot Prompting	89

3.1.3 Chain-of-Thought	93
3.1.4 Step by Step	96
3.1.5 Self-Consistency	99
3.1.6 Self-Refine	101
3.1.7 Tree of Thoughts <small>上級者向け</small>	105
3.1.8 Graph of Thoughts <small>上級者向け</small>	110
3.1.9 Dynamic Few-shot Prompting	121
3.1.10 ReAct	125
3.2 適切なプロンプト手法の選び方	132
第3章のまとめ プロンプトエンジニアリング	136

第4章

生成AIに新たな知識を与えるRAG 139

4.1 RAGって何？	140
4.1.1 RAGの基本要素とメリット	140
4.1.2 RAGが効果的なケース	142
4.2 実は検索エンジンが主役！RAGを支える仕組み	144
4.2.1 検索エンジンの基本用語	144
4.2.2 キーワード検索	147
4.2.3 転置インデックス	151
4.2.4 アナライザー	153
4.2.5 ベクトル検索	157
4.2.6 キーワード検索とベクトル検索のちがい	161
第4章のまとめ RAGの基本用語	168

第5章

実践！	
RAGシステムの導入ポイント	171
5.1 RAGシステムの全体像を知ろう！	172
5.2 「インデックスの登録」におけるポイントは？	174
5.2.1 「インデックスの登録」とは	174
5.2.2 レイアウト情報やメタデータを抽出する	176
5.2.3 チャンクの分割単位を見直す	178
5.2.4 想定質問を作成しておく	181
5.3 「ユーザーの入力」におけるポイントは？	182
5.3.1 「ユーザーの入力」とは	182
5.3.2 サンプルプロンプト集を表示する	184
5.3.3 メタデータで検索対象を絞り込ませる	185
5.3.4 ユーザーの属性にパーソナライズする	186
5.4 「検索クエリの実行」におけるポイントは？	187
5.4.1 「検索クエリの実行」とは？	187
5.4.2 LLMに検索クエリを生成させる	189
5.4.3 同義語辞書を使う	192
5.4.4 検索結果をリランキングする	193
5.4.5 ハイブリッド検索を用いる	194
COLUMN RRFの計算方法と基本的な性質	195
5.5 「回答の生成」におけるポイントは？	195
5.5.1 「回答の生成」とは	196
5.5.2 RAGシステムのユーザー体験を設計する	197
5.5.3 文書の参考箇所を提示させる	198
5.5.4 回答できない場合は次のアクションを示す	200
第5章のまとめ RAGの改善アプローチ	202

第6章

今エンジニアが最優先で学ぶべき技術！	
AIエージェント	205
6.1 AIエージェントって何？	206
6.1.1 実はシンプル！たった2つの仕組み	206
6.1.2 AIエージェントはLLMと従来技術の組み合わせ	207
6.1.3 AIエージェントを支える4つの機能	211
6.2 Profiling機能：役割と目的を定める	213
6.2.1 Role：役割と人格を設定する	214
6.2.2 Perception：相手と環境を知覚する	215
6.2.3 Regulation：全体の動作を規制する	216
6.3 Planning機能：目的を達成する計画を立てる	217
6.3.1 Decomposition：細かいタスクに分解する	218
6.3.2 Orchestration：タスクの実行を統制する	219
6.3.3 Refinement：計画を途中で見直す	220
6.4 Action機能：計画にもとづいて行動する	221
6.4.1 Selection：具体的な手段を選択する	222
6.4.2 Execution：行動を実行する	224
6.4.3 Evaluation：行動の結果を評価する	225
6.5 Memory機能：情報を記録して活用する	226
6.5.1 Writing：行動の情報を記録する	227
6.5.2 Reading：過去の情報を取得する	229
6.5.3 Reflection：過去の情報を内省する	230
第6章のまとめ AIエージェントの基本要素	232

第7章

実践！	
AIエージェントの応用テクニック大全	235
7.1 AIエージェント同士の連携パターン	236
7.1.1 単一型エージェント	236
7.1.2 階層型エージェント	238
7.1.3 ピアツーピア型エージェント	241
7.1.4 チーム型エージェント	244
7.1.5 社会型エージェント	247
7.1.6 マルチエージェントの構成のまとめ	250
7.2 AIエージェントの身近な例	252
7.2.1 コンピューター操作エージェント	253
7.2.2 調査エージェント	255
7.2.3 開発支援エージェント	258
7.2.4 汎用エージェント	264
7.3 AIエージェントを作るツール	266
7.3.1 Function Calling	266
7.3.2 AIエージェントフレームワーク	268
7.3.3 ノーコードツール	272
7.4 標準化ツール	274
7.4.1 MCP	274
COLUMN MCPはAIにとっての「USB-Cポート」	277
7.4.2 A2A	278
7.4.3 LiteLLM <small>上級者向け</small>	281
第7章のまとめ AIエージェントの応用テクニック	284

第8章

生成AIを「使いこなす」アクション集	287
8.1 生成AIマインドに切り替えるためのアクション	288
8.1.1 「精度100%の幻想」を捨てる	288
8.1.2 自分の言葉で「仮説検証」の数をこなす	291
8.2 プロンプトをチューニングするためのアクション	294
8.2.1 プロンプト改善の「基本フロー」を覚える	295
8.2.2 インターネット上のコンテンツに「スタイル」を近づける	297
8.2.3 「公式ガイド」に目を通す	299
8.2.4 オリジナルの「Few-shotサンプル」を作る	301
8.3 生成AIとの役割分担を見直すためのアクション	303
8.3.1 生成AIの「本質的な得意と苦手」を覚える	304
8.3.2 正しい情報に「グラウンディング」させる	307
8.3.3 「短絡的なタスク」を小出しに解かせる	309
8.4 生成AIの嘘を見つけやすくするためのアクション	312
8.4.1 「根拠の箇所」を抜粋させる	313
8.4.2 同じ質問を複数回試して「再現性」を見る	314
8.4.3 複数の情報源を「アンサンブル」させる	316
8.4.4 さまざまな形式言語で「可視化」させる	317
8.5 生成AIを感覚で覚えるためのアクション	321
8.5.1 1日1回「Deep Research」を使う	321
8.5.2 ChatGPTの「有料プラン」に1ヵ月加入する	323
8.5.3 わざと「アンチパターン」を試す	327
第8章のまとめ 生成AIを「使いこなす」アクション集	330

第9章

生成AIを「システム導入する」アクション集 333

9.1	企画・構想フェーズのアクション	334
9.1.1	目標精度は「70%」を目安にする	335
9.1.2	システムの目的を「汎用型」と「特化型」に分ける	337
9.1.3	「To-Beの業務フロー図」で自動化の範囲を描く	339
9.1.4	プロンプトの手動入力で「フィジビリティ」を確認する	342
9.2	精度分析・改善フェーズのアクション	344
9.2.1	正答率ではなく「業務効果」を定量化する	345
9.2.2	「定量分析」と「定性分析」の両方を行う	347
9.2.3	精度の「ボトルネック」を見極める	349
9.2.4	「Precision」か「Recall」か選ぶ	351
9.3	実機化・運用フェーズのアクション	353
9.3.1	「ウォーターフォール開発」か「アジャイル開発」か決める	354
9.3.2	「システム開発」にも生成AIを活用する	356
9.3.3	「継続的に改善する」アーキテクチャを設計する	360
	上級者向け	
	COLUMN アーキテクチャを設計するためには英語力も重要!	362
9.3.4	「OWASP Top 10 for LLM Applications」を読む	365
	上級者向け	
第9章	のまとめ 生成AIをシステム導入するアクション集	370
	おわりに	372
	索引	376
	著者プロフィール	385