


REV.	DRAWN BY:	DATE:
1	AI ARCHITECT	OCT 25, 2024



BLUEPRINT TO REALITY

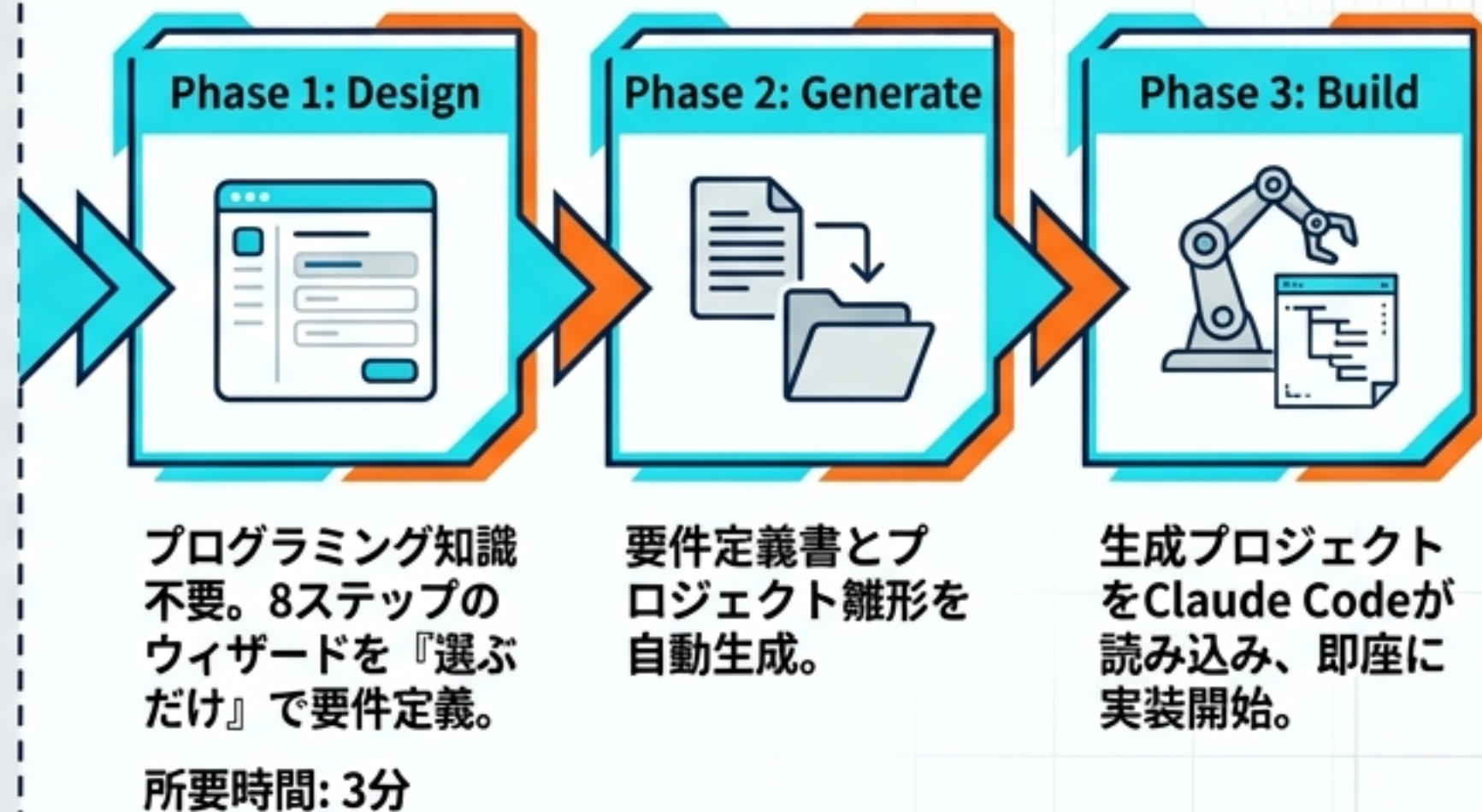
魔法のようにRAGを構築する設計ガイド

3分で要件を定義し、Claude Codeに実装を委ねる。
RAG Builderを利用した次世代の開発体験。

 SYSTEM INITIALIZATION : OK

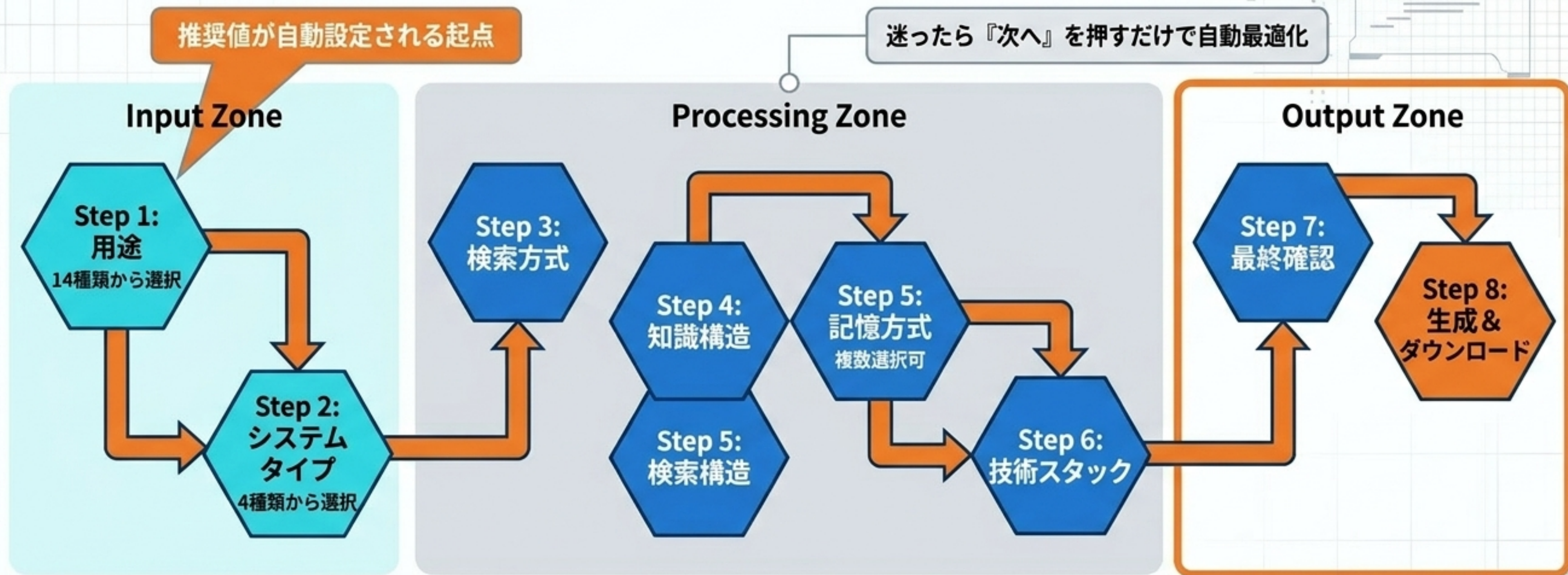
RES:	LIST00	06445	05TE	PROJECT: RAG SYSTEM CONSTRUCTION	BH
				DRAWN BY: AI ARCHITECT	
				DATE: OCT 26, 2024	

プログラミングから「設計とAI協働」へのパラダイムシフト



あなたは設計図を描くだけ。組み立てはAIが担当します。

The Pipeline: 8つのステップで要件定義書を生成する



あなたは設計図を描くだけ。組み立てはAIが担当します。

Phase 1: システムの目的と振る舞いを定義する

Step 1: 用途

14種類の用途（チャット系 / 検索系 / エージェント系 / その他）

[FAQチャットボット]

[社内ナレッジ検索]

[秘書AI]

Step 2: システムタイプ



チャット型

対話形式で回答

おすすめ: FAQ, サポート



検索特化型

文書をランキング表示

おすすめ: コンプライアンス



エージェント型

タスクを自律的に実行

おすすめ: 秘書AI, データ分析



マルチエージェント型

複数AIが協調

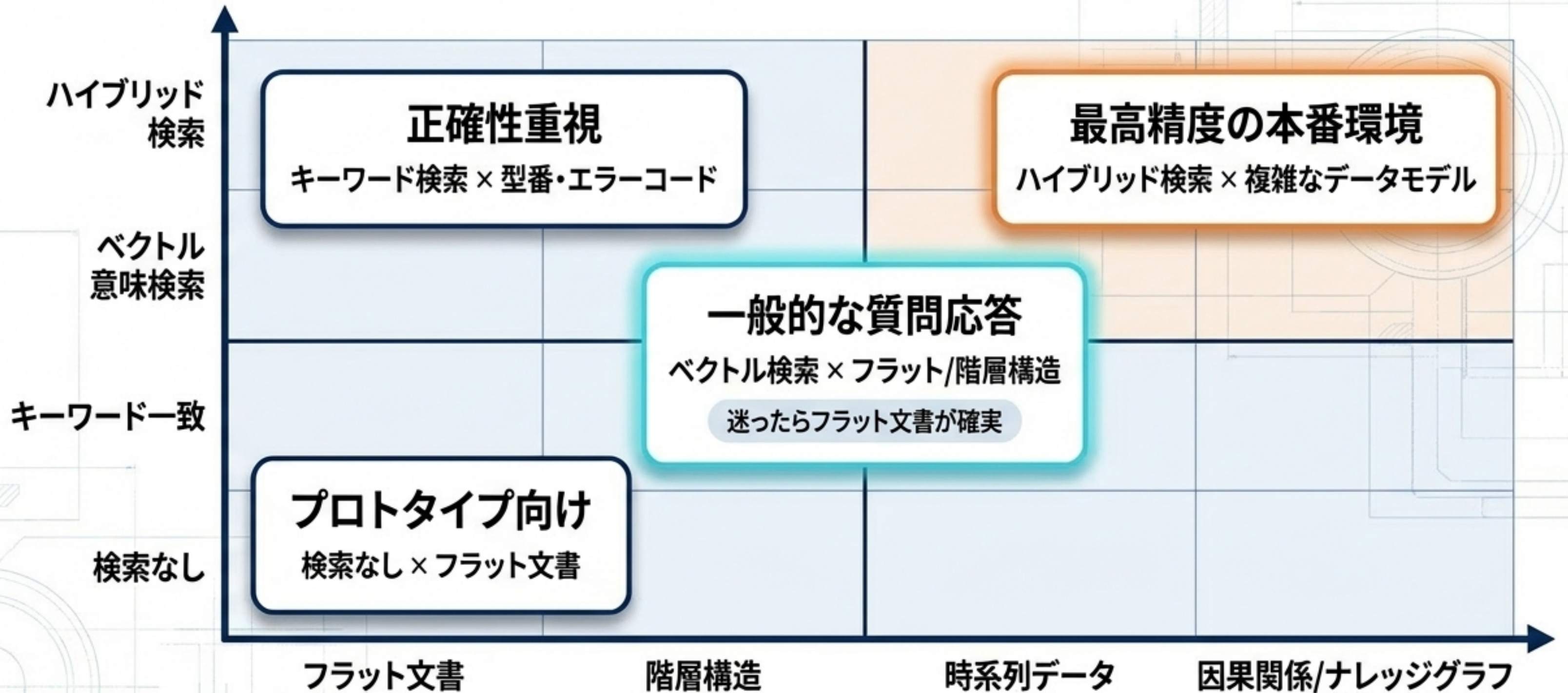
おすすめ: 複雑なリサーチ

! 『おすすめ』 バッジの選択肢が推奨ですが、カスタマイズも自由です。

あなたは設計図を描くだけ。組み立てはAIが担当します。

Phase 2: データへの最適なアプローチを決定する

Step 3: 検索方式 (精度と複雑さ)



注意: 『検索なし』を選ぶと、知識構造のステップはスキップされます。

Step 4: 知識構造 (データの複雑さ)

あなたは設計図を描くだけ。組み立てはAIが担当します。

Phase 3: システムの「記憶」と技術基盤を構築する

Step 5: 記憶方式



- 短期記憶**
今の会話の文脈を保持。
最近の会話を使用。※必須
- 中期記憶**
数日前の会話を覚えている。
- 長期記憶**
ユーザーの嗜好を長く記録して学習。
- プロンプト強化**
会話のヒントを使い、回答の品質と一貫性を向上。

Step 6: 技術スタック

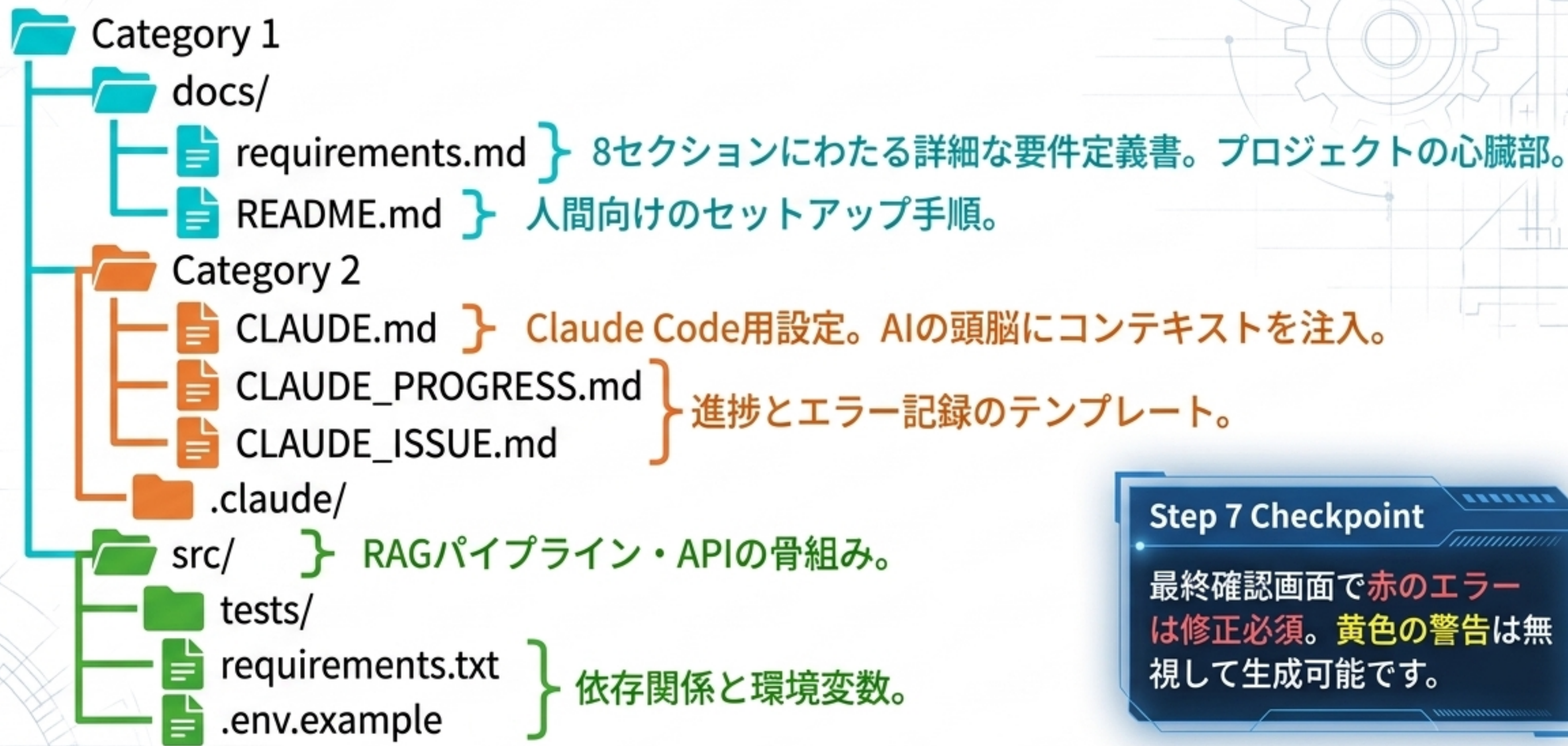


自動最適化基盤

用途選択時の推奨値が自動入力済み。
こだわりがなければそのまま『次へ』で完了。

あなたは設計図を描くだけ。組み立てはAIが担当します。

The Blueprint Output: 生成されるプロジェクトの解剖図

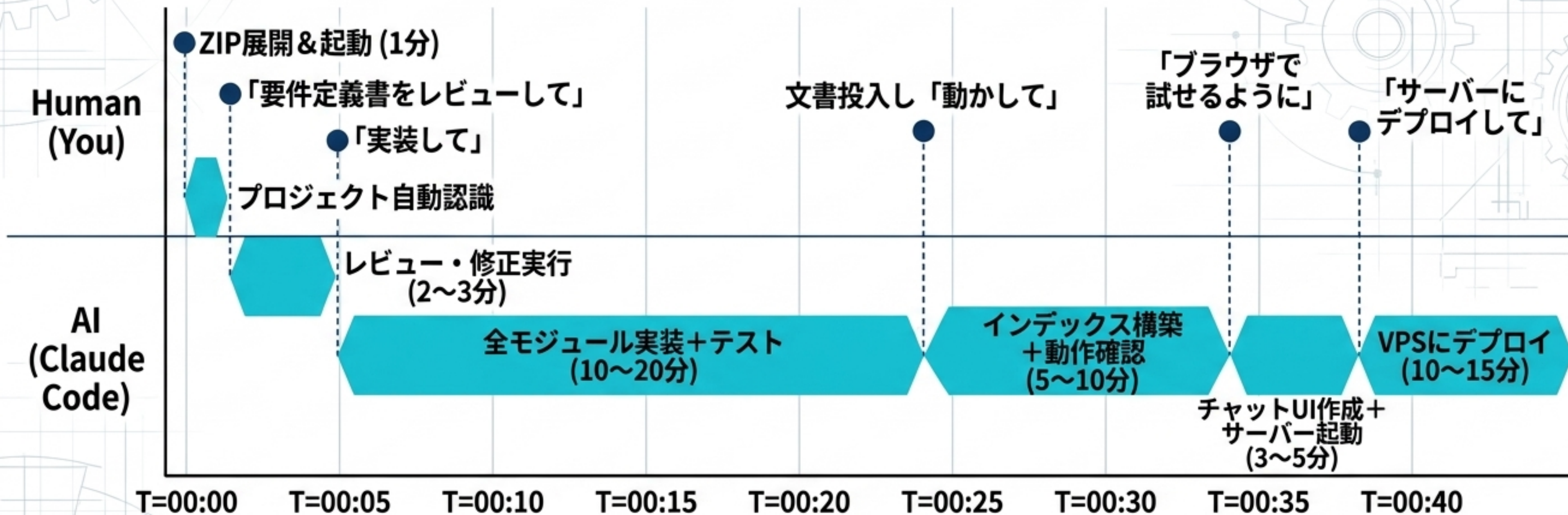


Step 7 Checkpoint

最終確認画面で赤のエラーは修正必須。黄色の警告は無視して生成可能です。

あなたは設計図を描くだけ。組み立てはAIが担当します。

Asymmetric Build: 人間とAIの非対称な協働タイムライン



Key Insight: 人間が1分のプロンプトを入力し、AIが10分働く。
あなたは『**作業者**』から『**指揮官**』へと昇格します。

Tactical Loadouts I: 目的別の最適な装備セット

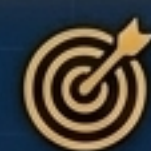


The Starter

初めてRAGを作る場合

最もシンプルで確実な構成。

用途 / タイプ	FAQチャットボット / チャット型
検索 / 知識	ベクトル検索 / フラット文書
記憶	短期記憶
Core Stack	OpenAI Embedding Chroma (Vector DB) Claude (LLM)



The Sniper

精度を最大化したい場合

検索漏れを防ぎ、最高品質の回答を生成する構成。

検索方式	ハイブリッド検索
Vector DB	Weaviate または Pinecone
キーワード検索	Elasticsearch
記憶	短期記憶 + プロンプト強化

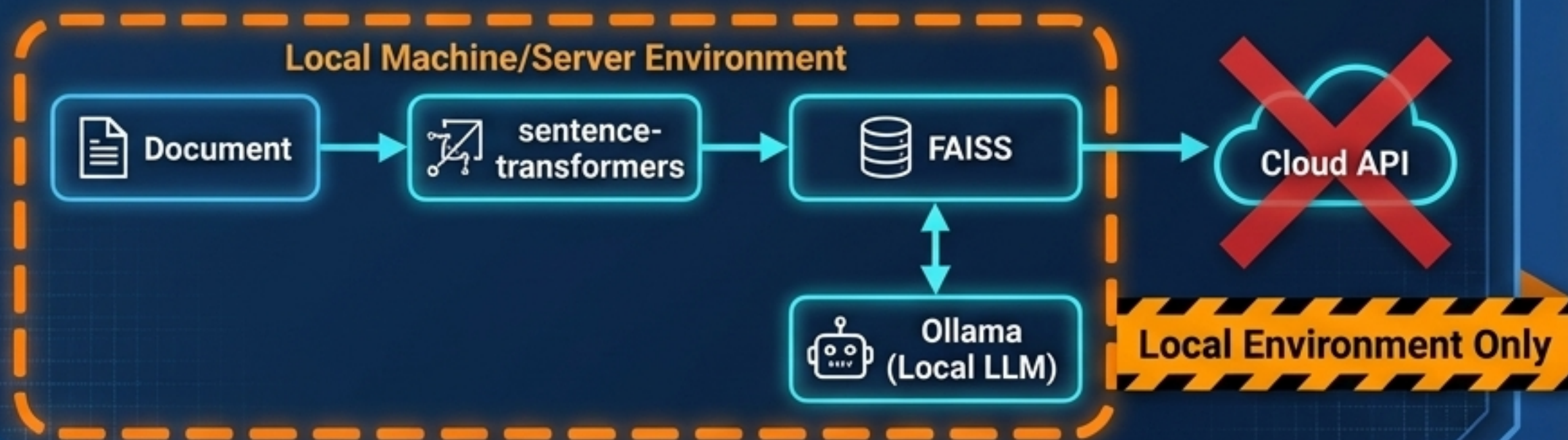
Tactical Loadouts II: 完全なる機密保持の実現

The Fortress

データを外部に出したくない場合

外部APIに依存せず、すべてのデータを自社環境内で完結させる構成。

Component	Implementation
Embedding	sentence-transformers (ローカル実行)
Vector DB	FAISS (ローカル・インメモリ)
LLM	ローカルLLM (Ollama経由で実行)



Diagnostic Flow: 問題解決とトラブルシューティング



ウィザード操作エラー

Symptom: 「次へ」ボタンが押せない。

Resolution: 選択肢を1つ以上選ぶこと。特にStep 5（記憶方式）は複数選択可能ですが「最低1つ」の選択が必須です。



検証エラー

Symptom: 最終確認画面で赤色エラーが出る。

Resolution: 例：「検索特化型には検索方式が必須」。
メッセージに従い該当ステップに戻り修正してください。



生成プロセスのフリーズ

Symptom: 生成に失敗する、ZIPがダウンロードできない。

Resolution: ブラウザをリロードしてください。解決しない場合はLocalStorageをクリアします。

```
F12 → Console → localStorage.removeItem('rag-builder:wizard-state') → Enter → F5
```

Architecture Reference: RAG構成要素の定義

