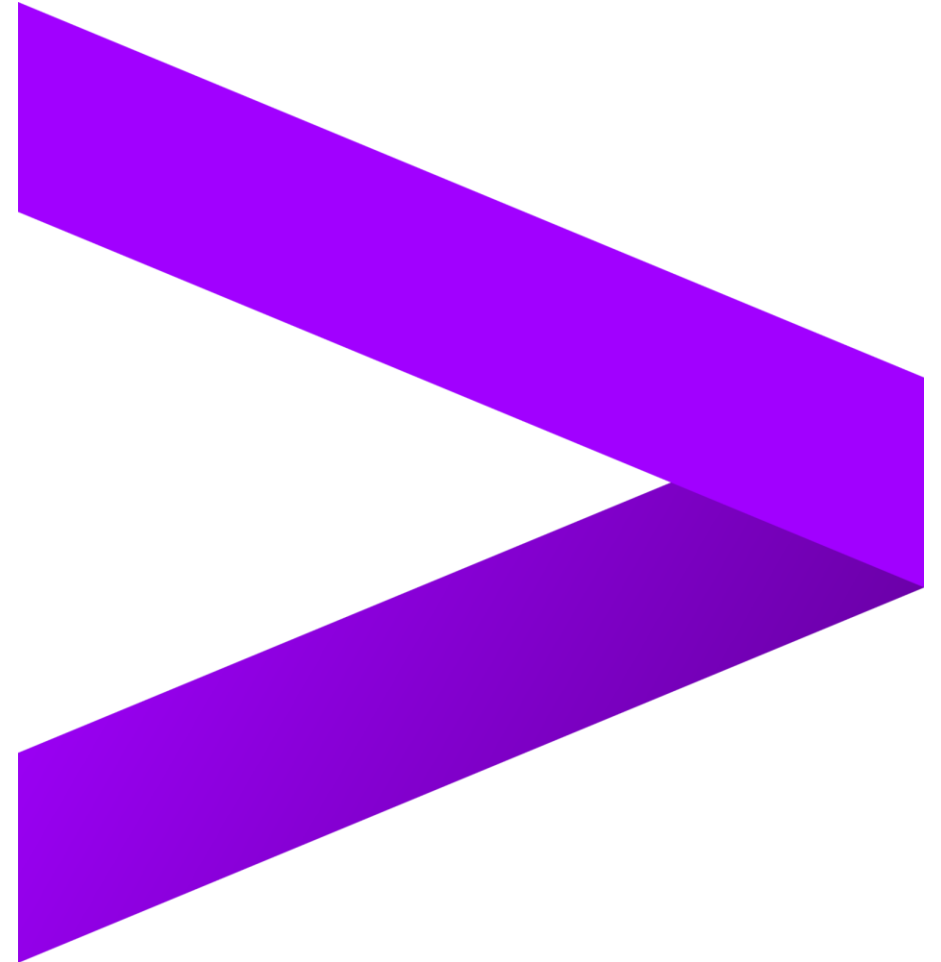


対話型AIを活用した 業務システム向けテストデータ 自動生成の実践的検証

2026/3/13

アクセンチュア株式会社 高橋 昭光



自己紹介



テクノロジーコンサルティング本部
Quality Engineering & Assurance

アソシエイトディレクター

高橋 昭光

略歴

20年以上にわたり金融系を中心としたテクノロジーコンサルティング、システム導入に従事

現在は**Quality Engineering**のリードであり、テスト・品質の専門家として大規模プロジェクトの**テスト戦略策定**／**品質評価**／**テストマネージャー**を担当

専門領域

- Quality Engineering
- プロジェクトマネジメント

業界経験

- 銀行/その他金融
- 損害保険
- 官公庁
- エネルギー



本日のロードマップ

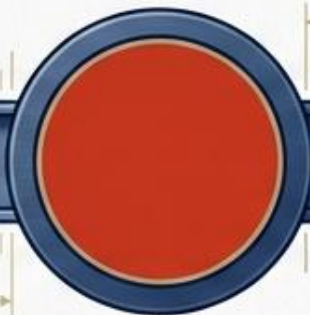
1



なぜ今、AIか？

現状の限界と新たなアプローチ

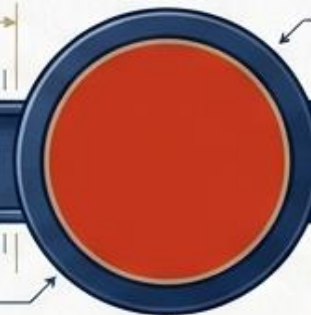
2



140万件の生成実験

LLMとの協働と3段階戦略

3



現場での評価と展望

圧倒的な成果と実務への適用

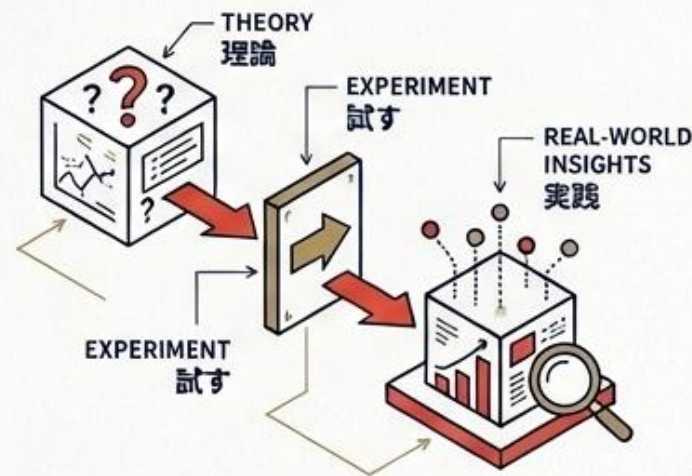
--- 章扉 ---

なぜ今、AIで テストデータ生成なのか？

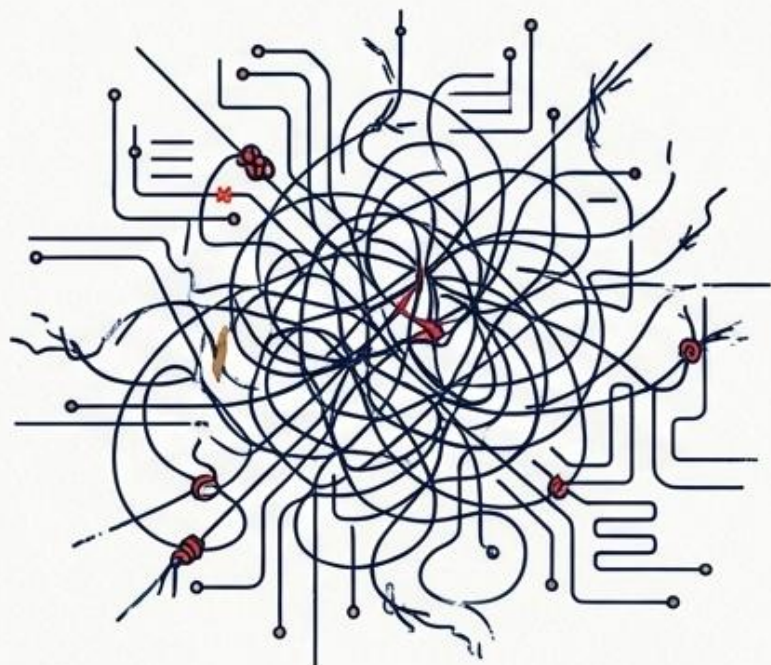


理論より、まずは 「AI for QA を試す」

AI技術は日進月歩であり、今日の手
法は明日には古くなる可能性があ
ります。しかし、理論をこねるより
も「まずは現場でAI for QAを試す」
ことにこそ真の価値があります。
本日はその「実践」から得られたリ
アルな知見を共有します。



複雑化するシステムと厳格化する法規制の板挟み



手作業の限界

複雑なリレーション、多数の外部キー制約、ドメインルール。
人手による整合性確保は多大な工数を要し限界に。



本番データ流用の限界

GDPRや個人情報保護法による厳格な制約。
マスキングによる運用はセキュリティリスクと隣り合わせ。

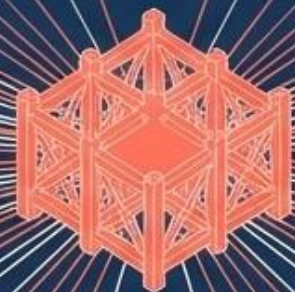
テストデータ生成手法のパラダイムシフト

手法	工数の低さ	法的クリア	構造整合性	分布制御
手作業	整合性確保に 多大工数	—	複雑システムで 困難	柔軟だが非効率
本番データマスキング	マスキングに 多大工数	個人情報保護法 ・GDPR等の制約	—	—
従来型自動生成ツール	—	—	複雑ルール 対応困難	緻密な分布制御 困難
提案手法 (LLMスクリプト生成)	✓	✓	✓	✓

--- 章扉 ---

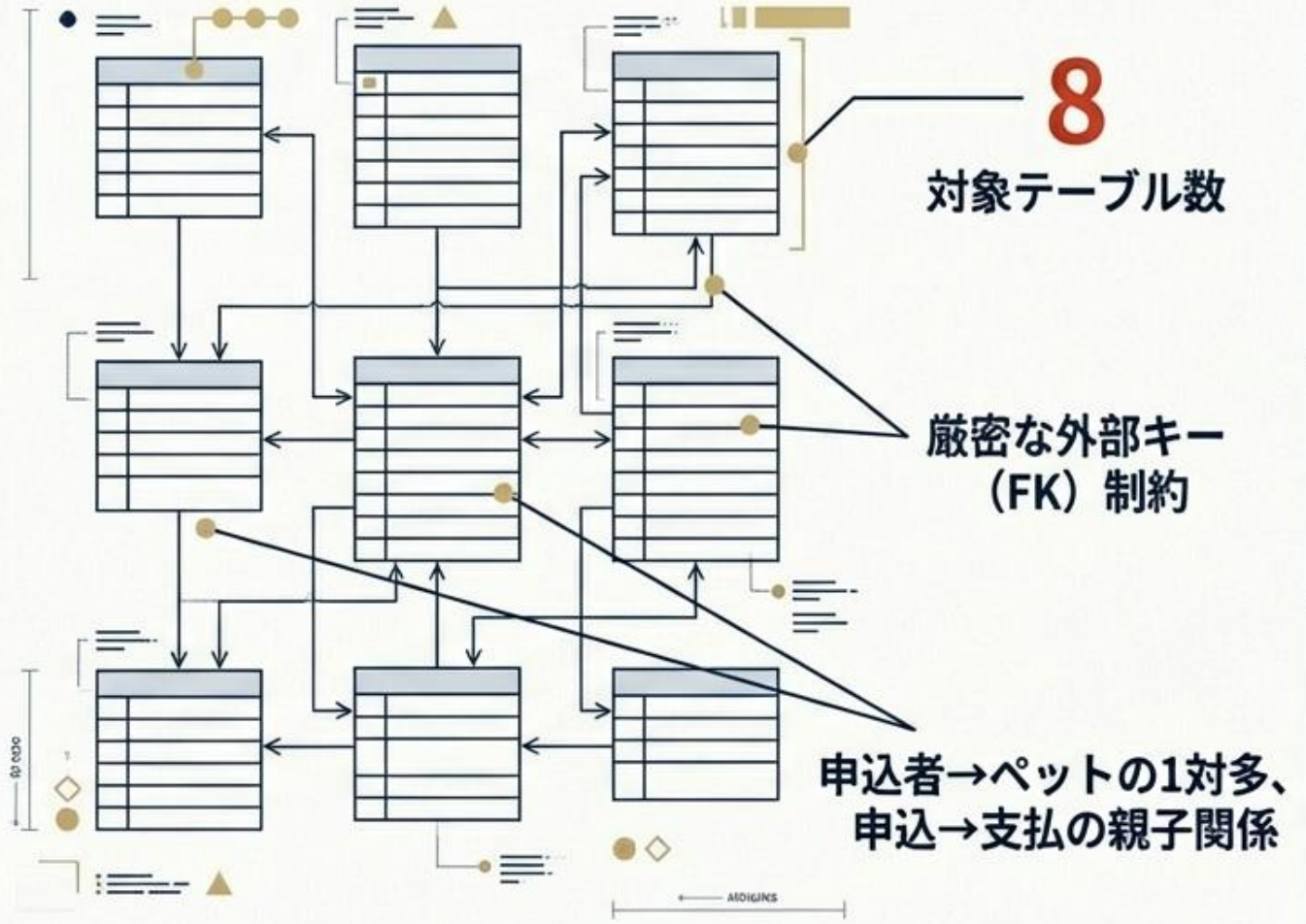
実際にやってみた:

140万件のデータ生成実験

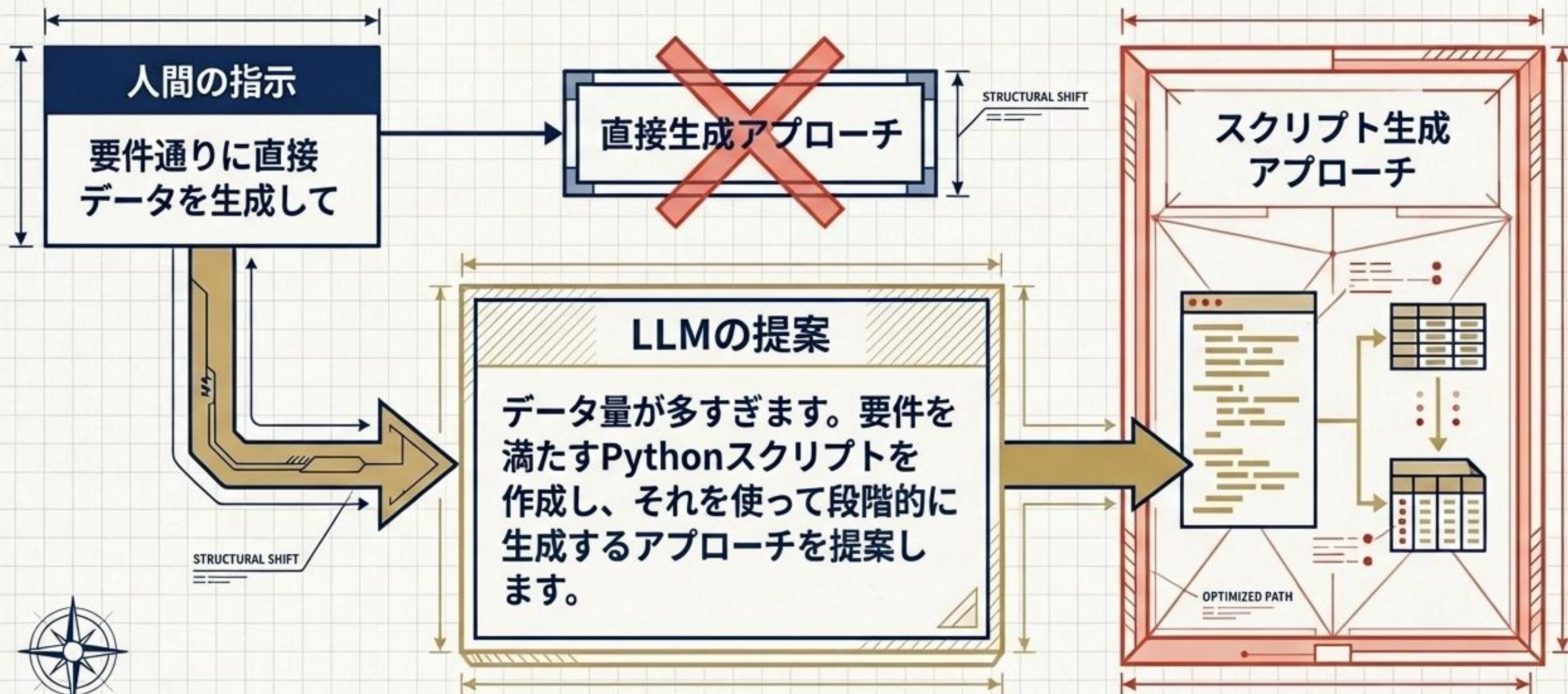


検証の舞台：複雑なリレーションを持つ架空のペット保険システム

1,400,000
総生成データ件数



LLMは「実行者」から「共創パートナー」へ



複雑性を紐解く「3段階生成戦略」



第3段階: 自然性向上

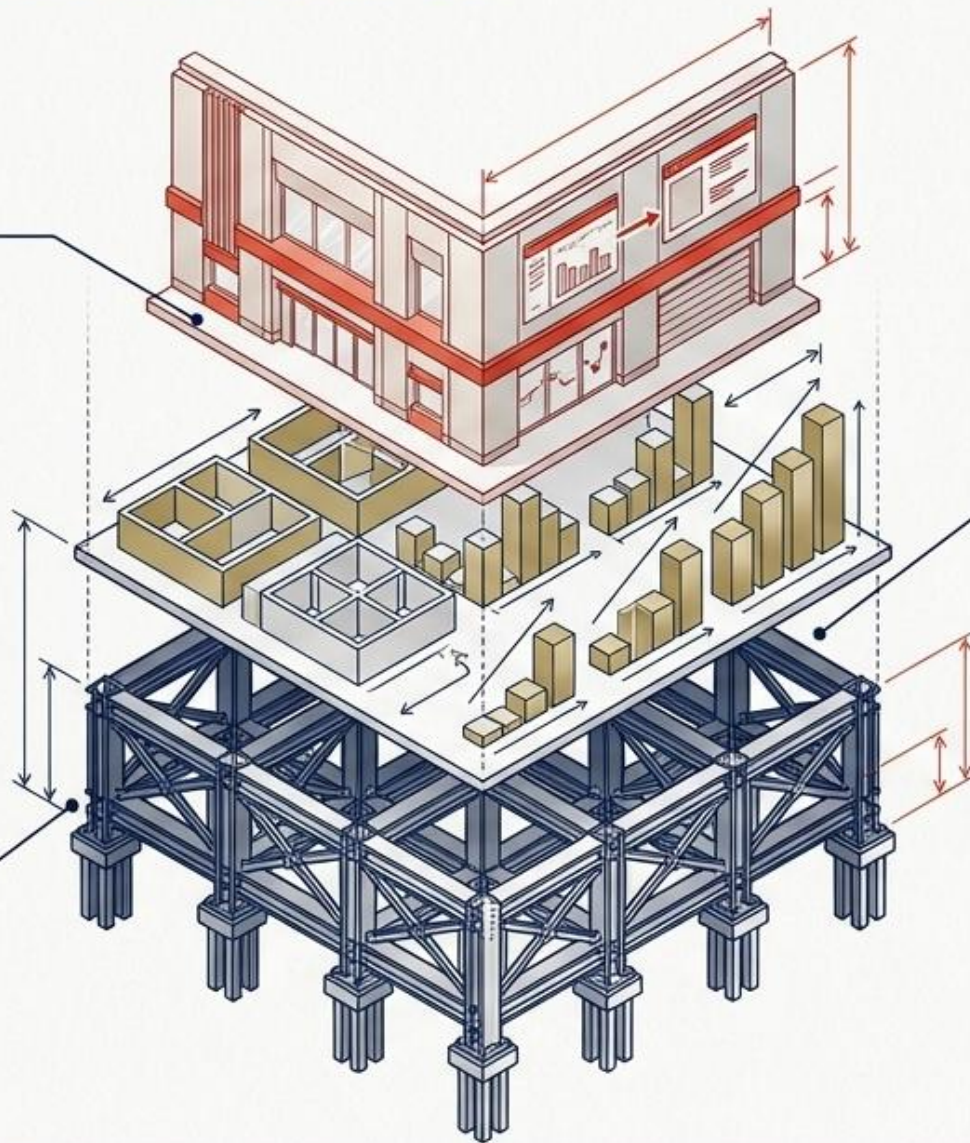
Faker等のライブラリ活用による、違和感のないリアルなダミーデータの肉付け。

第1段階: 構造整合性

ID一元管理と親→子への生成順序制御。違反率0%の骨格。

第2段階: 分布制御

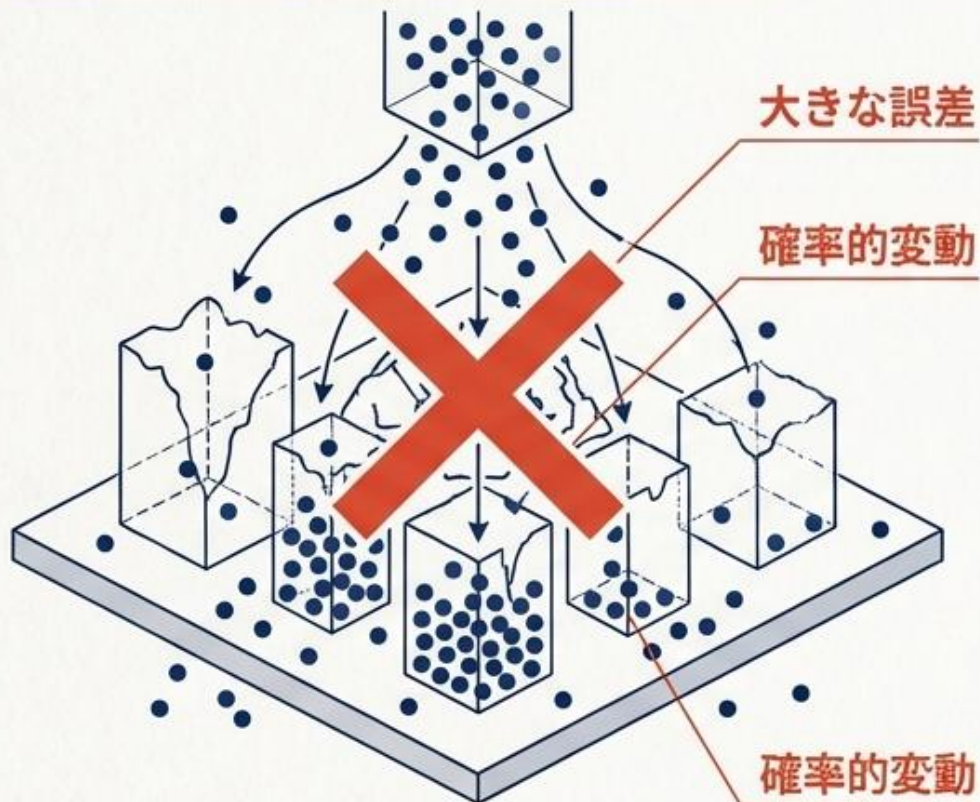
ビジネスのリアルな割合(年齢、地域など)の再現。事前割当方式の導入。



ランダム性の克服：LLMと共創した「事前割当方式」

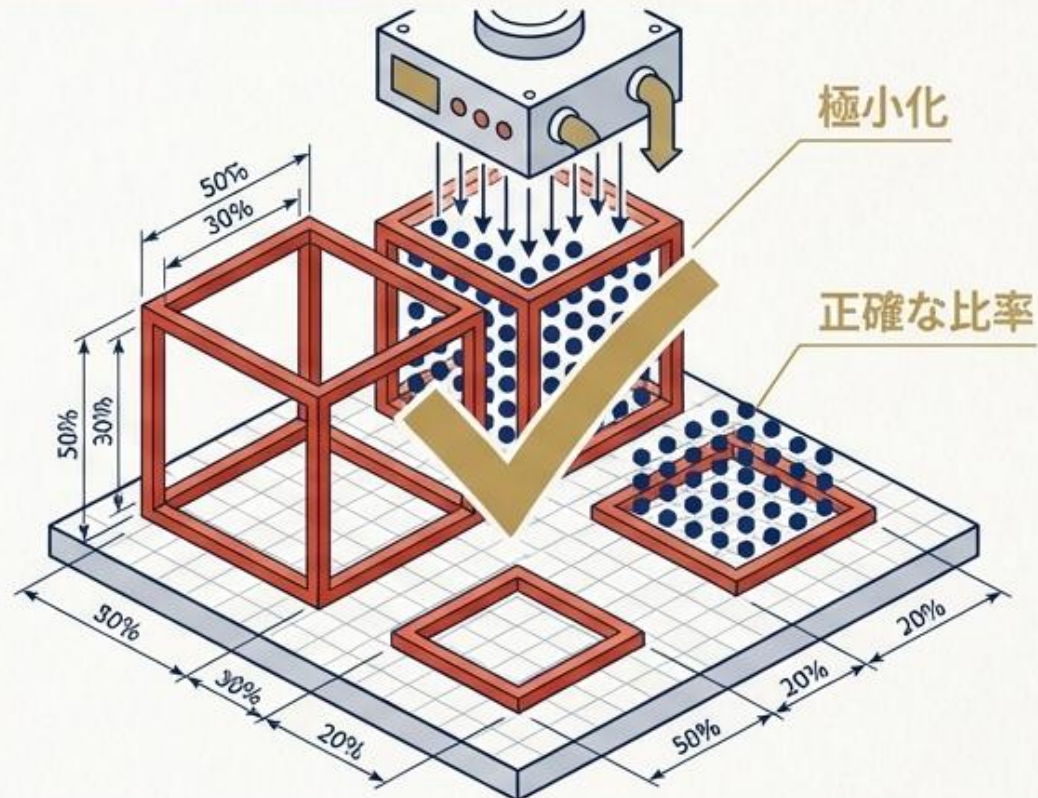


従来方式：ランダム生成→後分類



確率的変動による大きな誤差の発生

事前割当方式：枠の確保→逆算生成

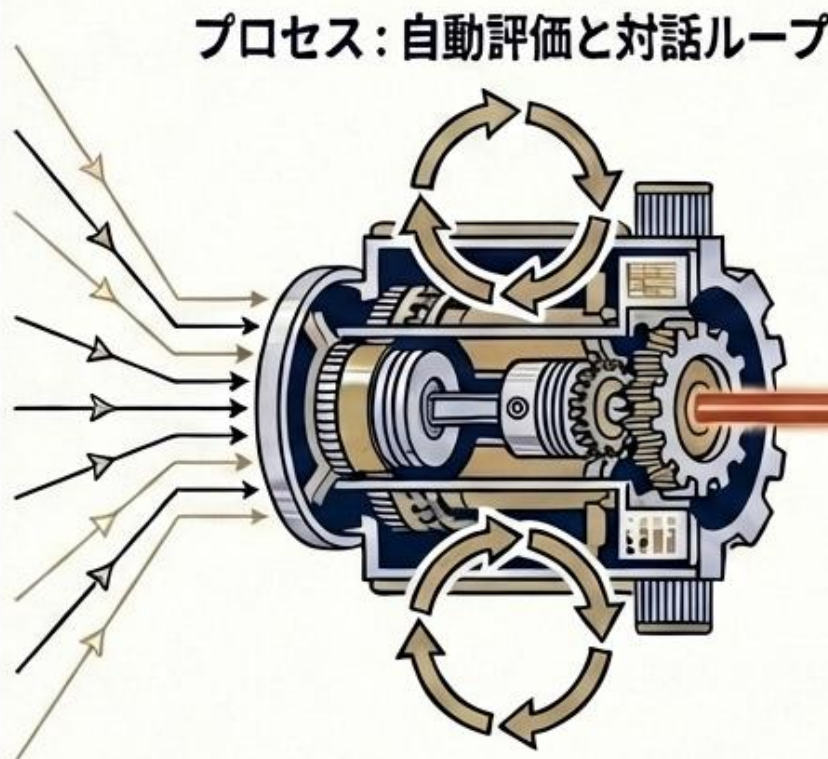


目標比率から枠を確保し、誤差を極小化

カオスから秩序へ：非決定性を制御するプロセス



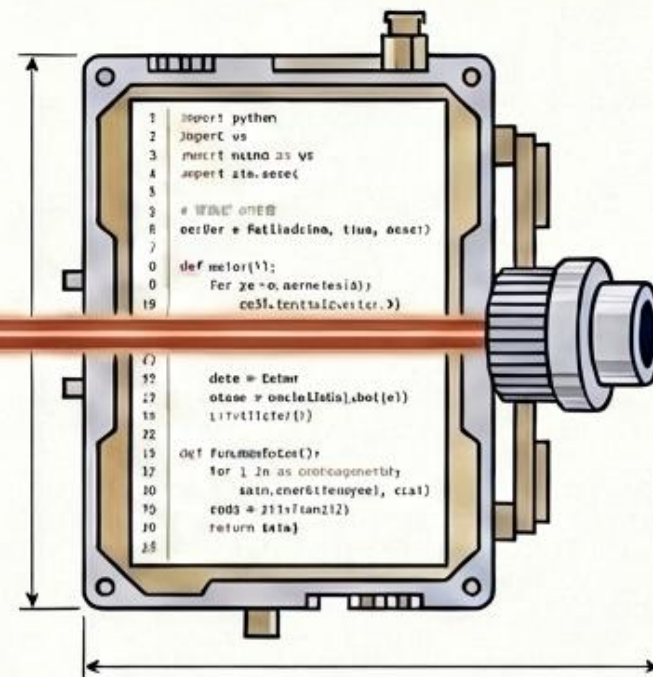
インプット：確立した標準構成



プロセス：自動評価と対話ループ

プロセス：自動評価と対話ループ

アウトプット：決定的な生成基盤



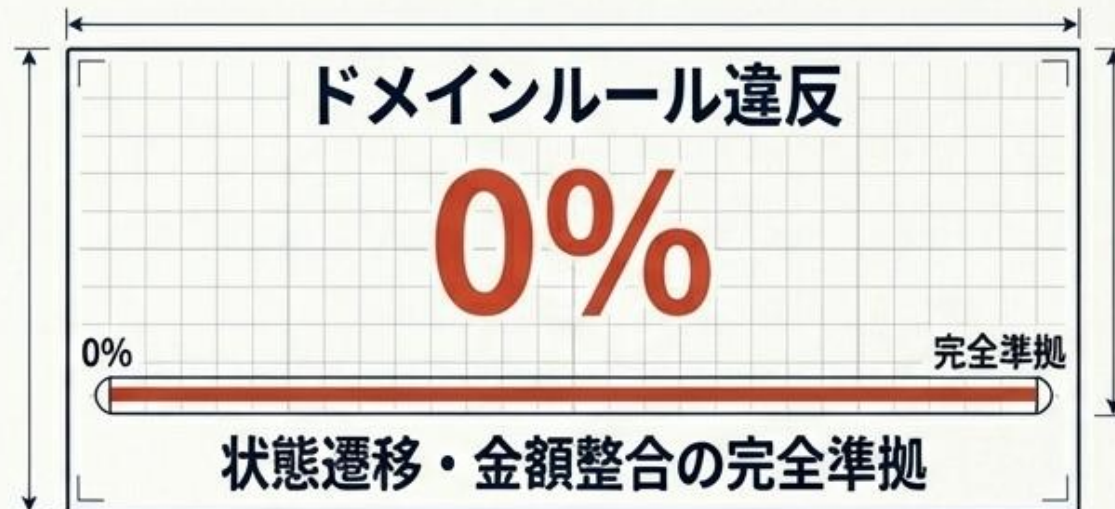
アウトプット：決定的な生成基盤
属人性を排除した高品質なPythonスクリプト

--- 章扉 ---

現場で使えるのか？

(評価結果と展望)

140万件生成における圧倒的な評価結果





実務適用における制約と向き合い方



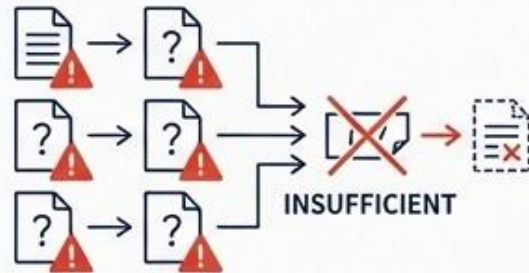
外部ライブラリの限界

Fakerの日本語氏名生成は内部辞書が少なく、**約25.5%の重複**が発生（カスタム辞書で対応可能）。



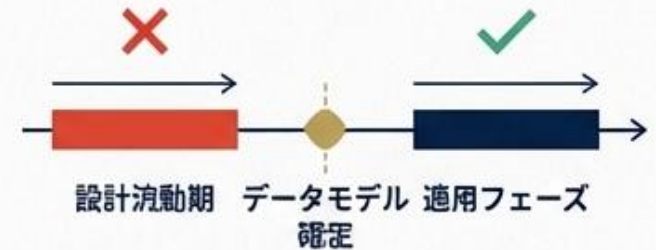
インプット品質への依存

9種類の標準資料や要件定義が不十分だと、生成結果の品質も担保できない。



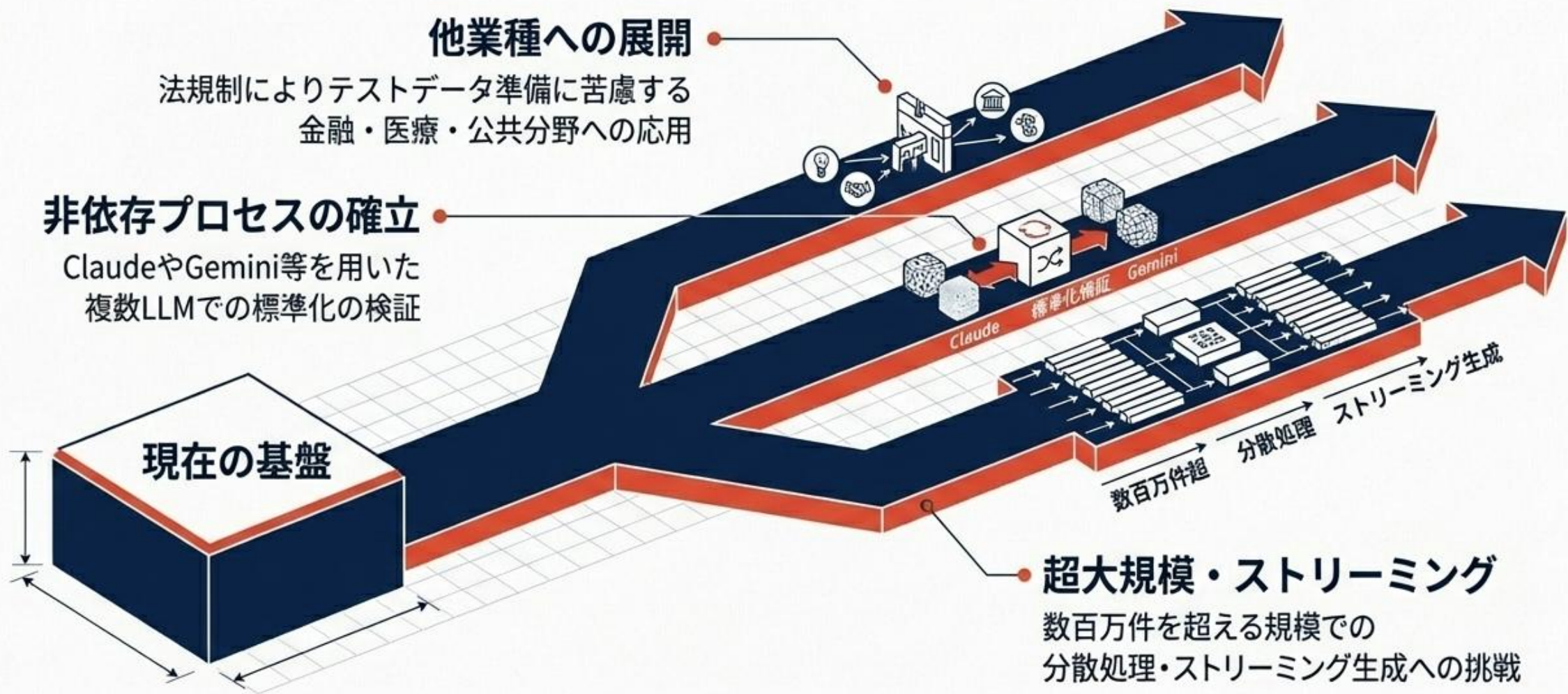
適用フェーズの見極め

設計流動期は改修コストが嵩むため不向き。**データモデル確定**後の適用が最適。

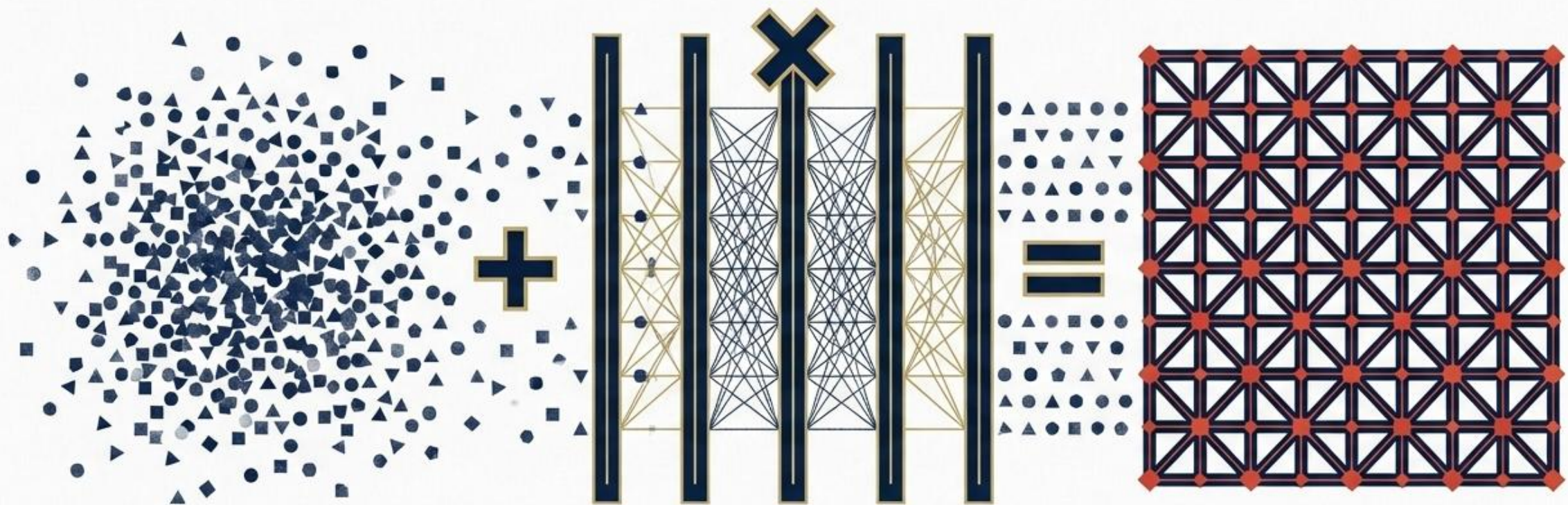




さらなるスケールアップと一般化へ



非決定的な技術を、現場で使える「武器」へ



Chaos

LLMの非決定性
+ 複雑な要件



Process

標準資料(9種) ×
段階的生成戦略



Structure

再現可能で決定的な
データ生成基盤

開発プロセスのパラダイムシフト

LLMはもはや単なる「実行者」ではなく、技術的課題をともに解決する「共創的パートナー」です。
テストデータの自動生成は、システム開発全体の生産性を底上げする確かな第一歩となります。