INDEPENDENT VERIFICATION REPORT

監査報告書：サマリア「特許調査・再現率100%」主張の検証と分析

生成AIによるAIスクリーニング（Top 30）の事実確認と、実務導入に向けた技術評価フレームワーク

対象サービス: サマリア（Summaria) / パテント・インテグレーション株式会社
分析対象: PR TIMES配信記事（2026年9月10日）
評価基準: データサイエンス、情報検索理論、特許調査実務

Gemini Notebook

事実：観測結果として整合

- PR TIMESに記載された「2024年度大会の既知正解特許26件を、社内AIベンチマークで上位30件以内に含めた」という事実は、公開情報と整合している。
- 明白な矛盾はない。

限界：普遍的性能の証明ではない

- これを「未知の特許調査一般で常常に再現率100%を出せる」科学的証明と読むのは誤り。
- 本結果はベンダーの自己評価であり、第三者監査や、ホールドアウトされた未見データでの実証を欠いている。

結論: 単純な「PRの鵜呑み」も「全面的な否定」も不適切。本作は、有望な「トリアージ・システム」としての真の価値と、適切な社内検証手順を解き明かす。

Gemini Notebook

検証の対象：PR TIMESにおける「100%」の定義

「上位30件を確認するだけで全分野が100%に到達」

電気	機械	化学・医薬
正解 4/4 件	正解 8/8 件	正解 14/14 件

合計 26/26件 (Recall@30 = 100%)

注：評価対象は3課題・既知正解文献26件に限定。

Gemini Notebook

主張と事実のギャップ診断（Verification Scorecard）

主張	判定	理由
会社が「26/26、上位30件で100%」と発表した	✔ 確認済み	PR TIMESと公式マニュアルの記載が完全に一致。
2024年度大会の実問題を使用した	⚠ 高い蓋然性	過去問の存在は確認。同社は「問2を使用」と具体的に説明。
大会公式基準で「上位30件100%」を達成した	✖ 支持されない	大会公式に「30件カットオフ」の基準は存在しない。ベンダーの独自指標。
独立第三者によって100%が検証済み	? 未確認	実行ログや外部監査の結果は公開されていない。
未知の特許調査でも100%を期待できる	✖ 根拠不足	サンプルが26件のみ。同一過去問での再テスト（後述）。
実務上有望な検索技術である	✔ 合理的に示唆	語彙拡張による大幅な改善（2月→9月）が観測されている。

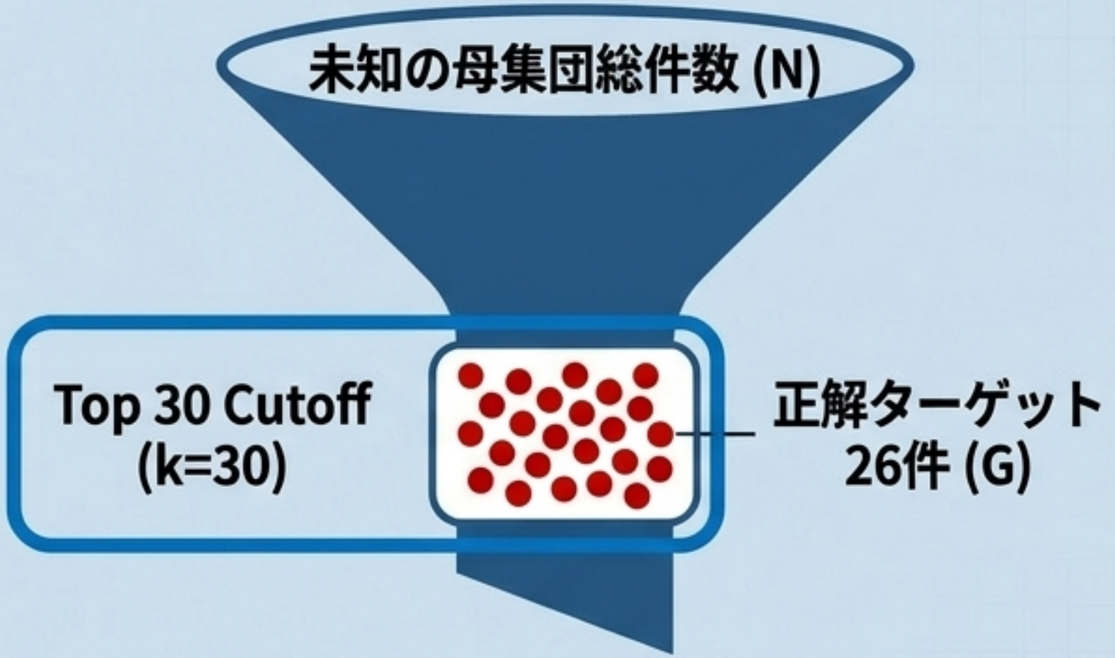
Gemini Notebook

指標の解像度：「公式大会の満点」ではなく「ベンダー独自指標 (Recall@30)」

IPCC公式の評価方式

- 概念: 再現率 (Recall) の全体評価。
- 仕組み: 競技者が作成した「ヒットリスト（検索集合全体）」を提出し、そこに含まれる正解文献の割合を採点する。
- 事実: 大会公式ルールには「上位30件に入れば合格」というカットオフ規定は存在しない。

サマリアの評価方式：Recall@30



- 仕組み: AIがスクリーニングした後のランキング上位30件に、26件の正解がすべて入っていたという「AIランキングの効率」を示すベンダー独自指標。

Gemini Notebook

「100%」の裏側：同一テストセットの再利用（A/Bテストの限界）

2026年2月（初期テスト）

同じ2024年問2を実施。機械 7/8件、化学・医薬 12/14件で「取りこぼし」が発生。

開発フェーズ

失敗箇所を把握。「類語シソーラス拡張」機能を新規実装。

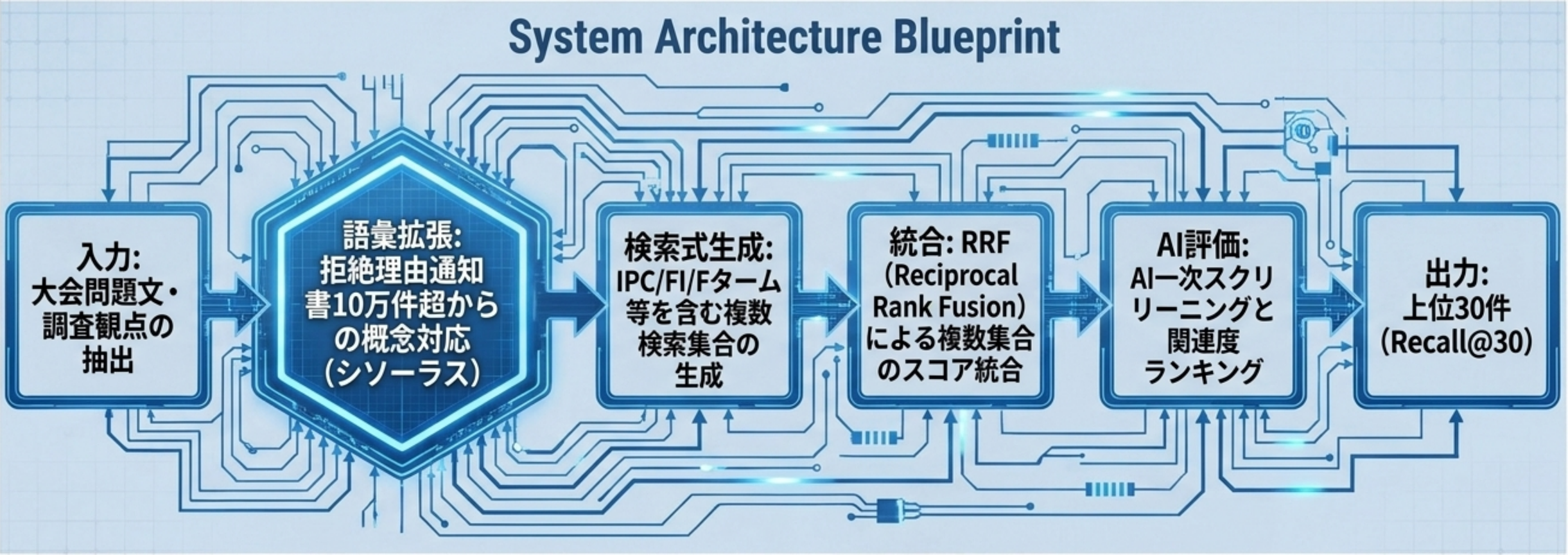
2026年9月（再テスト）

同じ過去問に再適用し、26/26（100%）を達成。

データサイエンスの視点: 製品改善のA/Bテストとしては極めて優秀。しかし、事前に失敗を把握した問題に対する再挑戦であるため、独立した「未知データへのホールドアウト・テスト」としては成立しない。過学習（適応バイアス）の懸念が残る。

Gemini Notebook

サマリアの技術的アーキテクチャ（単なるLLMラッパーではない）



技術的評価: 単一のLLMに依存せず、特許分類・キーワード検索・審査官由来シソーラス・生成AIによる意味評価を組み合わせた高度な「ハイブリッド型検索」である。

Gemini Notebook

コア・イノベーション：「審査官認定」の類語シソーラス

課題: 異なる専門語彙（出願人 vs 先行技術）が特許検索の最大の障壁。

解決策: 過去の審査官が「本願と引用文献の概念が対応する」と認定した事実（拒絶理由）を解析。

効果: 最大15件の類似審査事例を参照し、検索要素からAIスクリーニングまで語彙を引き継ぐ。2月の取りこぼしを埋めた中核技術。

Gemini Notebook

監査の死角：統計的妥当性を阻む「未公開データ」

1. 母集団の総件数（N）が不明

会社自身も未取得と明記。全体の候補数が分からないため、ランダムランキング時の偶然確率や適合率（Precision）が計算不能。

2. LLMのバージョンと安定性

OpenAI, Claude, Geminiを利用可能だが、100%達成時にどのモデルを使用したか、乱数（Seed/Temperature）による順位変動があるか不明。

3. シソーラスの学習データ（Cutoff）

10万件の学習データに、今回の評価対象文献や近縁のファミリー案件が含まれていないか（データ汚染の排除）が検証不可。

4. プロンプトと検索式の完全ログ

人手による介入がゼロであったことを第三者が完全に再現するための詳細ログが未公開。

Gemini Notebook

競合他社AIとの比較における「評価条件の不一致」

比較条件	サマリア （9月版）	GPT-5.2 / Gemini 3.0 （2月時点）
評価指標	Recall@30（上位30件）	設問(5)・最大10件抽出
評価時期	2026年9月（シソーラス実装後）	2026年2月
結果	100%（26/26件）	0%（抽出できず）
特許専用アーキテクチャ	あり（ハイブリッド型）	なし（汎用LLM）

分析：比較広告としての成立不可。評価のカットオフ（30件 vs 10件）が異なり、汎用LLMは特許のコンプリートサーチ専用製品ではない。
この0%をもって「汎用LLMは特許調査に無力」とは一般化できない。

Gemini Notebook

評価バイアスと限界に関する7つのチェックリスト

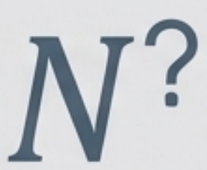
自己評価バイアス:
開発元自身による試験設計と実行。独立査読なし。

テストセット再利用:
失敗を知る同一問題での再テスト (既知への適応)。

データ汚染リスク:
学習データからの評価対象（ファミリー含む）除外手順が未確認。

極小サンプル:
3分野・計26正解文献のみ。統計的有意性なし。

非公式指標:
大会採点基準ではない、ベンダー独自指標（Recall@30）の使用。



母集団（N）不明:
Precisionやランダムベースベースラインが算出不能。

モデル安定性:
生成AI特有のRun-to-run（反復）分散の未検証。

注：会社自身もPR内で「小規模・母集団未取得」等の自己限定を明記しており、誇大広告というより「見出しの独り歩き」に注意が必要。

Gemini Notebook

実務的価値の再定義：「完全無欠の神託」から「超高効率のトリアージ」へ

従来の人手プロセス（100件の特許を全文査読）

約5時間の工数
（1件3分換算）

サマリア AIトリアージ（AIスクリーニング後の上位30件のみ査読）

約1.5時間（工数70%削減）

結論：「未知の特許を100%見つける」証明ではないが、「重大文献が上位30件に濃縮される」という事実は実務上極めて強力。全件査読を省略できなくても、読む順序の最適化と候補の圧縮において絶大なROIをもたらす。

Gemini Notebook

企業向け・AI特許調査ツール検証プレイブック（Enterprise Evaluation Framework）

[ ]	未見データでのテスト（Unseen Data）: 過去問ではなく、サマリア開発側が絶対にアクセスできない自社過去調査案件（30〜100件）を使用する。
[ ]	厳密な時間分割（Time Cutoff）: シソーラス構築時期以降の案件のみを使用用し、データ漏洩（Data Leakage）を物理的に排除する。
[ ]	ファミリー・引用の除外: テスト対象の文献群と直接リンクする引用案件を学習側から除外できるか確認する。
[ ]	ブラインド評価: 検索実行者（AI操作）と、正解を採点する担当者を完全に分離する。
[ ]	多角的な指標計測: Recall@30だけでなく、Recall@10 / @50、Precision、複数回実行の安定性（標準偏差）を計測する。

Gemini Notebook

Final Verdict（総合判定）

“

『100%』という数字は虚偽ではない。
しかし、それは極めて限定された社内テストセットにおける
『Top 30の圧縮率』を示すものに過ぎない。

サマリアの真の価値は、魔法のような絶対的網羅性ではなく、
審査官の語彙をハックした高度なハイブリッド・アーキテクチャによる
『査読対象の劇的なトリアージ能力』にある。

自社の未見データによる厳格なブラインド検証をパスすれば、
実務プロセスを変革する強力な武器となる。

Independent Verification Report concluding analysis based on PR TIMES (Sept 10, 2026) and supplied technical documentation.

Gemini Notebook