

AI Production Readiness Quick Scan 報告書（サンプル）

対象システム	Discord 常駐型 LLM エージェント（社内呼称 lab-llm_chatbot）
対象リビジョン	bb53090（main、2026-08-26）
実施日	2026-09-01
調査方法	書面調査のみ（リポジトリ、文書、CI 設定の読解と実測）
実施者	OriTech（株式会社七夕研究所）

AI Product Quality Assessment

Sample Report

v1.0

0. この文書について

これは OriTech の成果物のサンプルです。対象は七夕研究所が自社で開発し、社外の利用者がいる環境で運用している実システムです。他社の顧客システムではありません。顧客システムの報告書は守秘義務により公開できないため、自社システムに同じ手順を適用したものを公開しています。

1. この文書がしないこと

Quick Scan は危険箇所を短期間で把握するための商品です。次は含みません。

- 出荷の可否を判定しません
- 判定のための証跡整備はしません。個々の指摘に確認方法は書きますが、それ以上の体系は作りません
- 本番環境に触っていません
- 実データを見ていません
- 原因からの横展開探索（同型の欠陥が他にないかの探索）はしていません

この文書だけを「出してよい根拠」として第三者に示すことはできません。それが必要な場合は Assessment（判定と証跡を作る商品）が対象になります。

2. 危険箇所

優先度順に 7 件。優先度は、壊れたときの影響と気付けにくさの掛け合わせで付けています。確認方法が書けないものは載せていません。

R-1 LLM の出力を検査する資産がない（優先度：高）

評価データセット、ゴールデンセット、自動判定のいずれも存在しません。追跡されている全ファイルを対象に、名称と内容語の両方から探索し、ヒットした候補は読んで確認した上での結論です。同じ探索方法で、存在すると分かっている資産（運用手順書やレビュー記録）が検出できることも確かめており、探索そのものが機能していないわけではありません。

このシステムの中核機能は LLM の応答そのものです。モデルの更新、プロンプトの変更、スキル定義の変更のいずれについても、応答の質が落ちたことを検出する手段がありません。モデルをティアで使い分ける設計になっていますが、ティア配分を変えたときの影響も事後に確認できません。

いま機能している唯一の検出手段は、利用者や開発者が使っていて気付くことです。これは利用者が少ないあいだしか保ちません。テストが 1000 件通っていることや行カバレッジが 91% であることは、この不在とは関係がありません。テストが検査しているのはコードの振る舞いで、応答の質そのものではないからです。

R-2 本番事象の記録が分散していて、一箇所から追跡できない（優先度：中）

実機で見つかった不具合の記録自体はあります。日付付きの報告書に「ドッグフーディングで見つけた不具合と修正」の節があり、コミットメッセージにも詳細が残っています。ただし統一した索引がありません。報告書、変更履歴、レビューの処置文書それぞれに事象の記述が散っていることは、内容を読んで確認しました。

事象から原因、そこからの横展開、回帰テストへという流れを一箇所からたどる経路がありません。個々の記録の質は悪くないのに、この事故から何を一般化し、どこまで手を広げたかを後から追う手段が欠けています。問題は記録の不在ではなく分散です。

R-3 仕様 ID とテストの対応を機械が検査していない（優先度：中）

仕様は REQ → FUNC → DET の 3 層で ID 体系を持っています。仕様 3 文書からユニークな ID を網羅的に抽出すると 256 件（REQ 89、FUNC 81、DET 86）。テストコードを検索して参照を数えると、ユニークで 69 件でした。

内訳の偏りがはっきりしています。69 件のうち 68 件が DET 層への参照で、REQ 層への参照は 0 件。テストが実際に触れているのは 3 層の一番下だけで、要求レベルの記述にはテストからのリンクが一本もありません。

対応関係の完全性を機械が検査する仕組みもありません。仕様を追加・変更したときに、対応するテストが無いことに気付く手段がないということです。なお、参照のない ID をテスト未実施とは述べていません。記述のみの仕様項目もあり、参照の有無から即座にテストの有無は導けません。だからこそ、意図的に検査対象外だと宣言できる仕組みが要ります。

R-4 CI に依存パッケージの脆弱性検査がない（優先度：中）

CI 定義で動いているのは lint とテストだけで、依存の更新を自動で検査する仕組みもコミット前のフックもありません。外部と通信する常駐プロセスである以上、依存の脆弱性が検出されないまま残る期間が生れます。組織の他リポジトリでは秘密情報スキャンが動いていますが、本リポジトリで別経路が用意されているかどうかは未確認のままです。

R-5 ブランチ保護がサーバ側で強制されていない（優先度：中）

主要ブランチへの直接反映を禁じる規約と、CI の可視ステータスだけで運用しています。使っているブランチの制約で強制設定を有効化できないと、README にそのまま書かれています。CI が緑になったのを確認してから反映する、という手順が人の遵守に依存している状態です。開発者自身がこの制約を README に明記している点は評価できます。隠れた不備ではなく、既知の制約として扱われています。

R-6 テストが観測していない範囲が宣言されていない（優先度：低）

テストを完走させて集計すると、行カバレッジは 91%（3672 文中 336 文が未実行）。悪くない水準ですが、この 336 文が意図した除外なのか単なる漏れなのか、区別する記録がありません。分岐カバレッジも測っていません。

R-7 テストの層が構成上区別できない（優先度：低）

契約テストと受け入れ試験は事故対応として後から足されていますが、ファイルの配置と命名からは契約や結合、受け入れといった層を見分けられません。どの層が薄いかを構成だけで読み取れない状態です。

3. 先に触れておくべき点

危険箇所だけを並べると実態を見誤ります。同規模のシステムとしては規律が高い方だと見ています。

- テストの厚み：1000 件が全て成功し、行カバレッジは 91%（3672 文中 336 文が未実行）
 - 独立レビューの記録：生成とは別系統の LLM によるレビュー記録が 12 件残っている。指摘と処置の対、あるいはそれに相当する統合文書を 1 セットとして数えた単位で、中には複数回のやり取りを含むものもある
 - 運用手順の文書化：デプロイ・再起動、初回デプロイ、講義後のデータ撤収という 3 つの手順が、実際に書かれた形で残っている
 - 秘密情報と個人情報：実値を除外し、テナント別の設定を暗号化してコミットしている。実名簿もパターンで除外され、追跡されていない
 - 権限境界：外部通信を行う補助コンポーネントをネットワーク分離していて、ホストへ公開していない
 - 事故からの学習：実機で見つかった不具合を個別修正で終わらせず、不変条件と回帰テストに変換している
 - 公開指標の再現性：外部記事が公開したテスト件数を独立に再計測して、一致を確認した
-

4. 確認していない範囲（簡易）

Quick Scan は書面調査のみです。次は見えていません。この範囲に問題がないことを意味しません。

- 稼働中の本番環境（コンテナの状態、権限の実体、外部サービスのクォータ）
 - プロンプトインジェクション耐性の実試験
 - 実データでの挙動
 - 監視のアラート設定と通知経路
 - 復旧手順の実演
 - 依存パッケージの脆弱性の実スキャン
 - 分岐カバレッジ
 - コードの論理的正しさ
-

5. 次に取り組むべき一手

最優先は R-1 です。他の 6 件は運用を止める理由にはなりませんが、R-1 は利用範囲を広げられない理由になります。

R-1 への着手は外部委託なしに、自社で始められます。手順は次の 3 つです。

1. 実際の会話ログから代表事例を 20～50 件抽出する
2. それぞれについて「期待される応答の性質」を言語化する。事実性、根拠の提示、指示遵守、文体、過剰な断定の回避。網羅性より、壊れたら困る事例が入っていることを優先する
3. モデルやプロンプトを変えたときに、その事例集で人が読み比べる運用にする

自動化は後で構いません。まず変えたときに読み比べる対象が存在することが重要です。

この先に判定が必要になる場合

次のいずれかに当てはまる場合、Quick Scan の範囲では足りません。

- 顧客や上司、監査に対して、出してよいと示す根拠が必要
- LLM の出力がそのまま外部に提示される用途へ広げる
- 個人情報・成績など、誤りが当事者に直接不利益を与えるデータを扱う
- テナント数を増やし、1 テナントの障害が他へ波及し得る構成にする

その場合は Assessment（判定と証跡を作る商品）が対象になります。本サンプルの Assessment 版は `2026-09-01-ai-production-readiness-assessment-sample.md` にあります。

6. 本報告書の限界

- 本報告書は保証ではありません。実施した書面調査の範囲における所見です
- 出荷判定ではありません（第 1 章）
- 所見は調査時点（bb53090）のもので
- OriTech は本システムの開発主体である七夕研究所と同一法人です。本サンプルは独立性の要件を満たしていません
- 本サンプルは公開用のため、構成の詳細を一般化して記述しています。確認に用いた具体的な手順は記載していません。実案件の報告書では、第三者が同じ結果にたどり着ける形で証跡として添付します