

AI エージェント向け「品質評価サービス」新規開始 ～振る舞いまで含めた品質を、評価観点に基づき第三者の立場で評価～

ソフトウェアの品質向上支援に関するサービスを提供する株式会社ベリサーブ（本社：東京都千代田区、代表取締役社長：嶋原 忠大、以下 ベリサーブ）は、あらゆる AI エージェント^{※1}の品質を評価する新サービス「QA4AI（キュー・エー・フォー・エーアイ）エージェント」の提供を本日より開始しました。

本サービスでは、AI エージェントの出力結果だけでなく、判断過程やツール利用といった振る舞いまで含めた品質を、評価観点と評価プログラム^{※2}に基づき第三者の立場で評価します。

※1 外部ツールやデータと連携し、業務やタスクを自律的に実行する AI

※2 評価指標（メトリクス）やスコアリング、判定基準などに基づき、AI エージェントの品質を評価するための仕組み

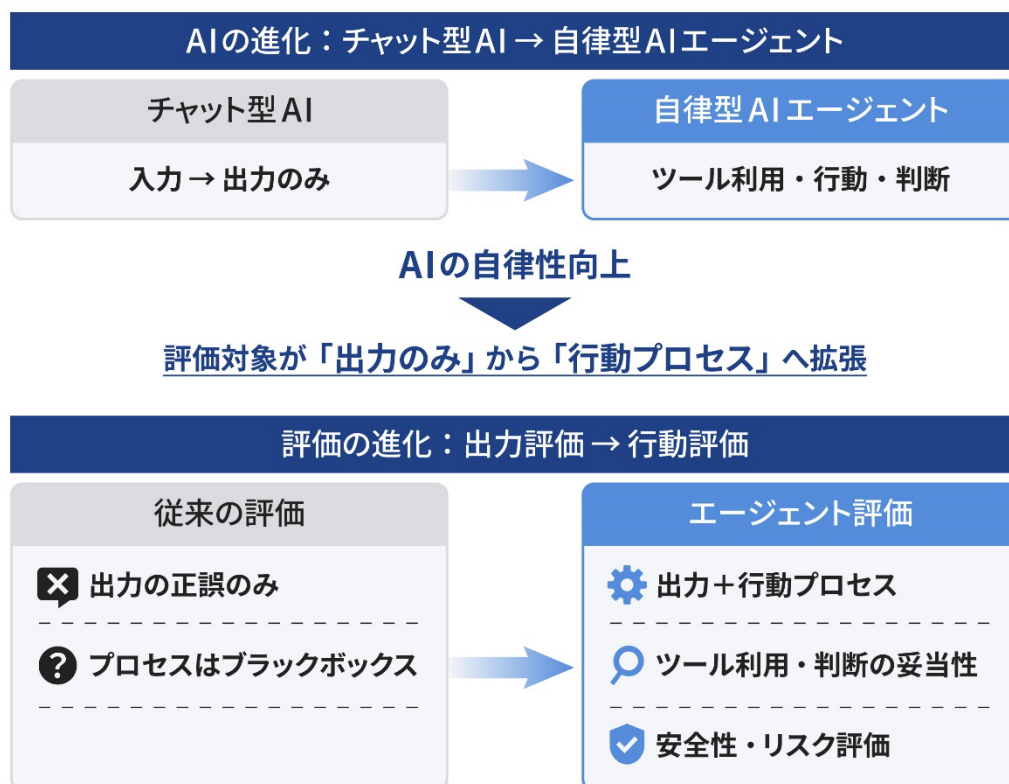


図1 「行動プロセス」も対象にするAI エージェント評価へ進化

■背景

生成AIの活用は、従来のチャットで出力を得ることから、人間が行う業務そのものを実行するAIエージェントへと広がっています。一方で、AIエージェントは複数のタスクを自律的に分解・実行し、途中結果に応じて次のアクションを選択する特性を持つため、最終的な出力結果だけでなく、タスク分解や実行過程、判断の妥当性を含めて評価しなければ品質を十分に担保することが難しいという課題が顕在化しています。

また、企業においても、AI エージェントの導入や本番利用を検討する中で、「どのように品質を評価すべきか分からない」「変更による影響を適切に把握できない」といった課題が見られています。

ベリサーブはこれらの課題に対し、AI エージェント特有の品質を評価する観点を体系化し、その観点に基づいた評価プログラムの適用からテスト実行までを一貫して提供することで、あらゆるAI エージェントの品質を客観的に評価し、企業が安心して導入から本番利用できるよう支援します（図1）。

■サービス概要

「QA4AI エージェント」は、AI エージェントを開発する企業や導入予定の企業に向けて、導入前や本番利用前、モデルや設定の変更時などに品質を継続して評価する新サービスです。

ベリサーブが第三者の立場で、評価設計から実行までを担い、AI エージェントの品質を客観的な観点に基づき評価します。

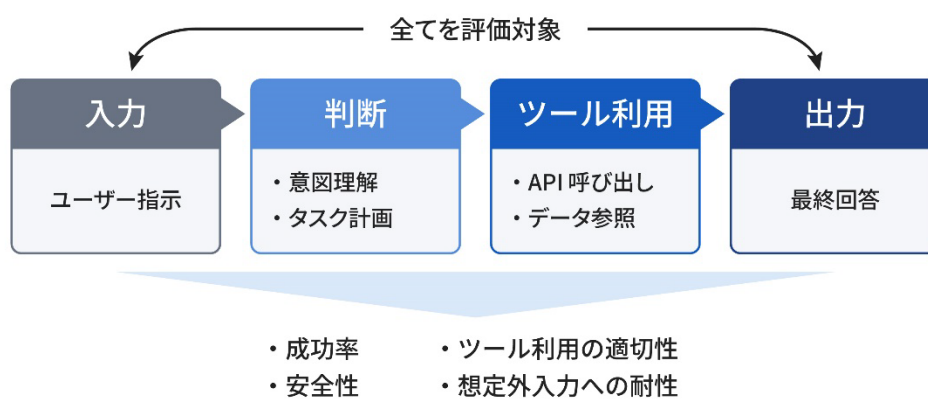
【主な支援内容】

- 本番利用前の品質評価および可視化
- リスクの特定と低減に向けた検証
- 仕様変更に伴うソフトウェア改変時の品質影響の評価
- 導入判断に必要な評価結果の提供

【主な実施内容】

- 現状分析とスコープ定義
- 評価観点の整理
- 評価指標（メトリクス）の設計
- データセット設計
- 評価スクリプトの実装および評価実行と結果の報告

AIエージェントは“どこを評価するのか”



AIの結果だけでなく、**判断や動きまで評価**

図2 AI エージェントの品質評価

■特長

1. 振る舞いまで含めた品質評価

従来のAI に対する評価観点は出力結果の正しさが中心でしたが、本サービスでは以下のような観点でAI エージェントの振る舞い全体を評価します（図2）。

- 意図した成果物が出力されているか
- タスクが適切に完了しているか
- ツール利用が適切か
- 権限外の情報にアクセスしていないか
- 想定外の入力にも安定して対応できるか
- 安全性・コンプライアンス上の問題がないか

2. 評価観点の体系化と評価ツールの活用

ベリサーブは、AI エージェント特有の品質を評価する観点を体系化すると共に、その観点に基づく評価ツールを整備しています。

これにより、異なる出力であっても同じ基準で品質を評価し続けることが可能となり、品質を一貫した基準で評価しソフトウェア改変時においても継続的に評価・比較できます。

3. 第三者による品質評価

ベリサーブは、長年の研究開発活動とさまざまな産業分野での豊富な実績に基づくソフトウェアの品質向上の知見を有しています。加えて、AI エージェント特有の評価観点や評価プログラムを活用し、第三者の立場で、開発者自身の確認だけでは見落とされがちなリスクの検出を支援します。

■具体的なユースケース

(1) AI エージェントを開発する企業

開発段階から第三者の品質評価を取り入れることで、利用前にリスクを可視化し、安心して本番利用できます。

- 本番利用前に、想定外入力や複雑なシナリオでの振る舞いを検証
- モデル変更、プロンプト変更、ツール追加時の品質影響を評価
- 出力の正しさだけでなく、ツール利用や判断の妥当性を確認

(2) AI エージェントを導入する企業

導入前に品質を客観的に評価することで、業務適用におけるリスクを低減し、安心して導入判断が可能になります。

- 業務シナリオに基づいた実運用に近い形での検証
- 誤回答や不適切な振る舞い、リスクの洗い出し
- 自社の業務要件に対する適合性の評価
- 導入可否や適用範囲の判断材料を提供

■今後の取り組み

ベリサーブでは、さらに進化していく AI エージェントに対応した品質保証に継続的に取り組んでいきます。評価手法の高度化を進めると共に、評価観点や評価プログラムの拡充を図り、「人×技術×AI」による新たな品質保証の形を実現し、お客様の安全・安心なソフトウェア開発に貢献していきます。

■株式会社ベリサーブについて

設立：2001年7月24日

代表者：代表取締役社長 嶋原 忠大

本社：東京都千代田区神田三崎町 3-1-16 神保町北東急ビル

事業内容：ソフトウェア事業

1. ソフトウェアテスト・品質関連事業
2. サイバーセキュリティ関連事業
3. コンサルティング関連事業
4. ソフトウェア開発関連事業
5. その他事業

URL : <https://www.veriserve.co.jp/>

【サービスに関するお問い合わせ先】

<https://www.veriserve.co.jp/contact/>

【リリースに関するお問い合わせ先】

広報部 広報課 佐藤、太田

TEL : 050-3640-8194

MAIL : press@veriserve.co.jp

※掲載されている製品名、会社名、サービス名は、各社の商標または登録商標です。