Citadel AI

AI 品質を評価・向上させる Citadel AI のご紹介

2024年3月
株式会社Citadel AI

AI の品質・信頼性に関わる新たな課題を解決

株式会社 Citadel AI

- 2020年12月　設立
- 2021年9月　Seed Funding
- 2023年6月　Series A

Citadel AI

総額 5.2 億円の Series A資金調達完了

UTokyoIPC　CORAL　anri　SUNTORY　MUFG Mitsubishi UFJ Capital

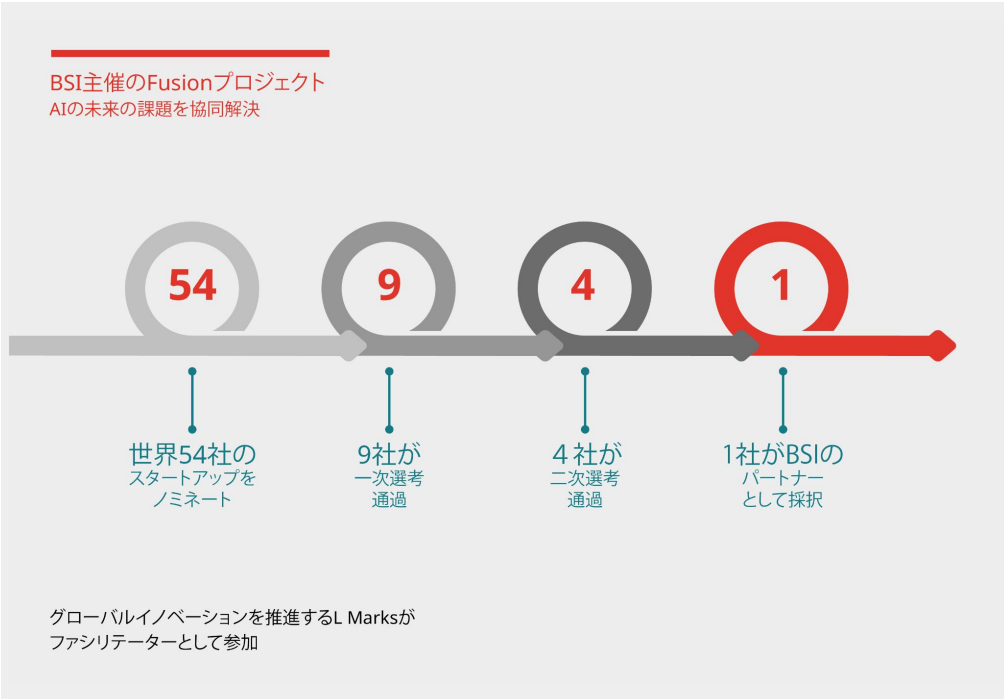
Google　WAYMO　paidy　TOYOTA　stripe

最新版 すごいベンチャー、スタートアップ 100 に選出

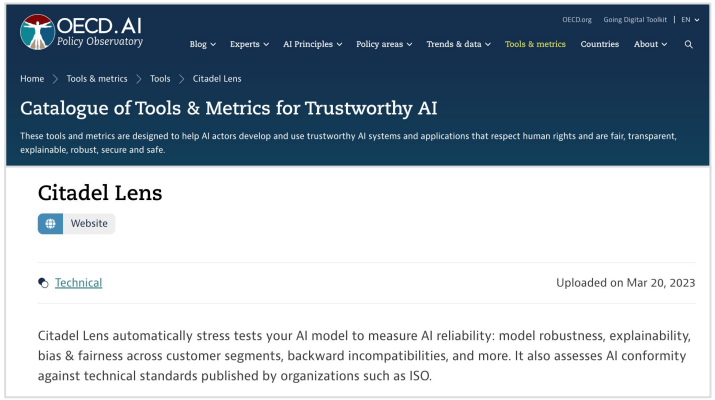
世界の AI 国際標準をリードする英国規格協会が AI の認証審査を念頭にグローバルに採用

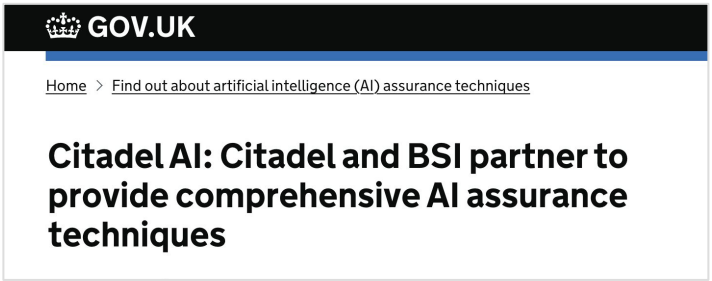
bsi.

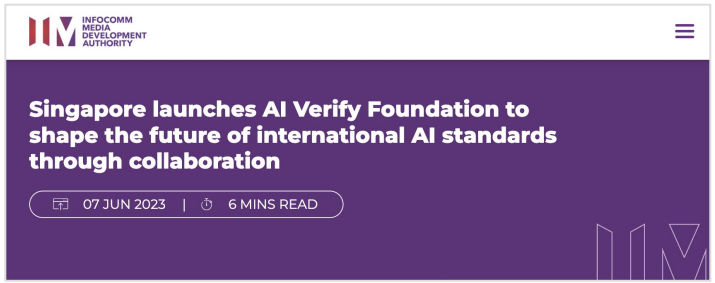
国際的な規格開発と
認証分野で世界をリードする
BSIにて正式採用決定



米国 AI Safety Institute 等 海外公的機関の取組みに参画







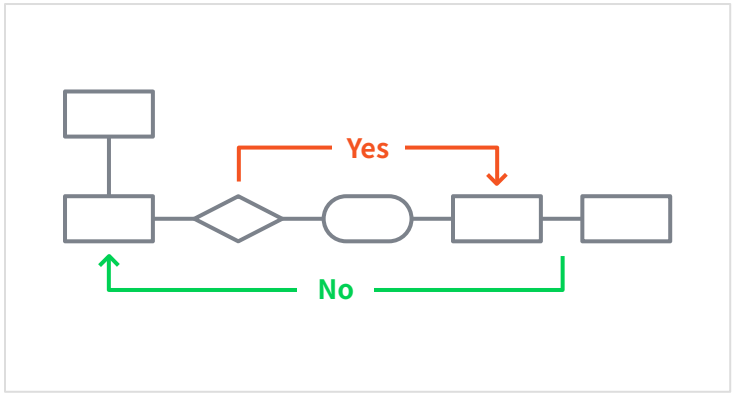
1

Citadel AI が考える
AI 品質に関わるリスク

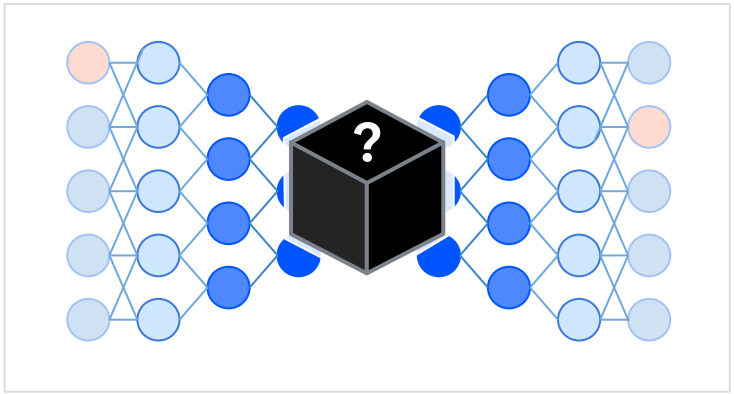
Citadel AI が考える、AI リスクの全体像

1 法的なリスク	各国の AI 法規制に対するコンプライアンス
	個人情報保護、知的財産権
2 社内体制上のリスク	AI ガバナンスの制度設計・体制づくり
	AI 人材の不足
3 技術的なリスク	AI 固有の新たな技術的リスク
	従来のシステムでもあったサイバーセキュリティ

AI はブラックボックス

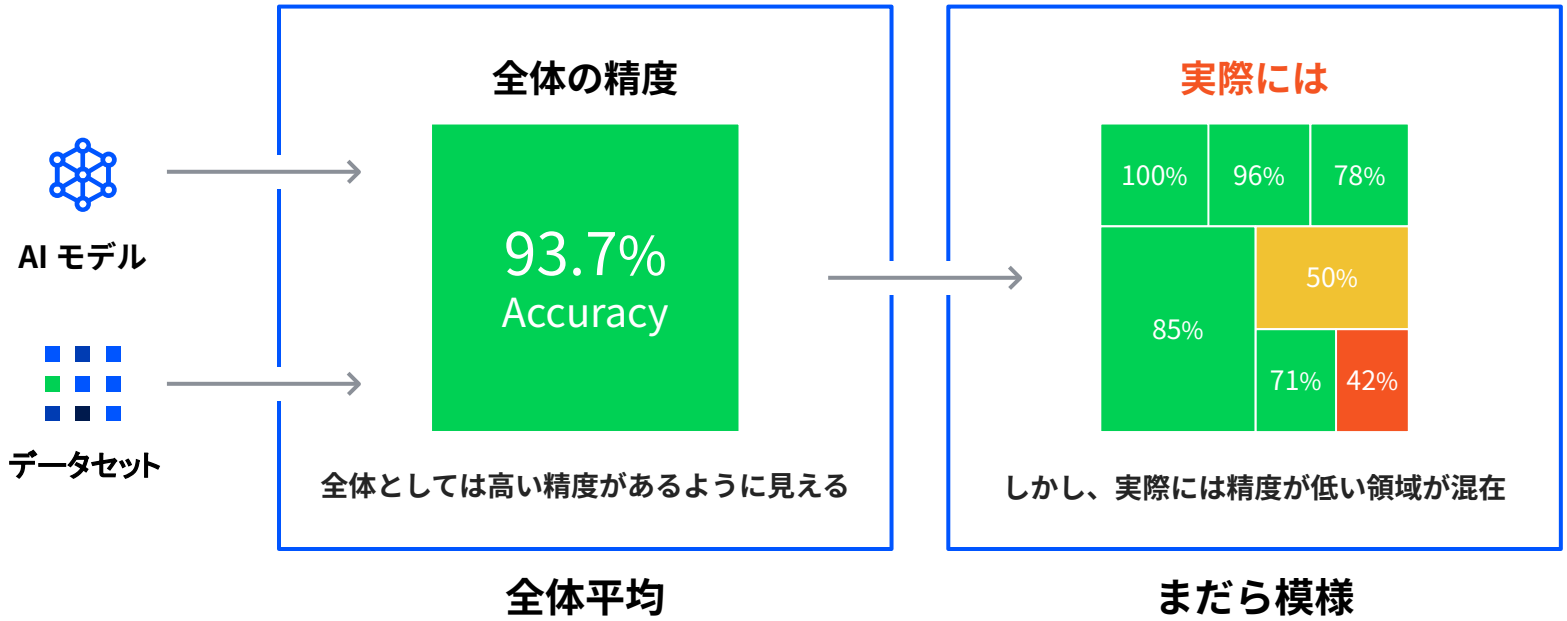


従来のソフトウェアの構造



AI ソフトウェアの構造

学習時 (開発時) の課題：隠れた未学習領域に残るリスク



運用時の課題：AI は環境変化に脆弱でドリフトを起こす

既存顧客

新規顧客

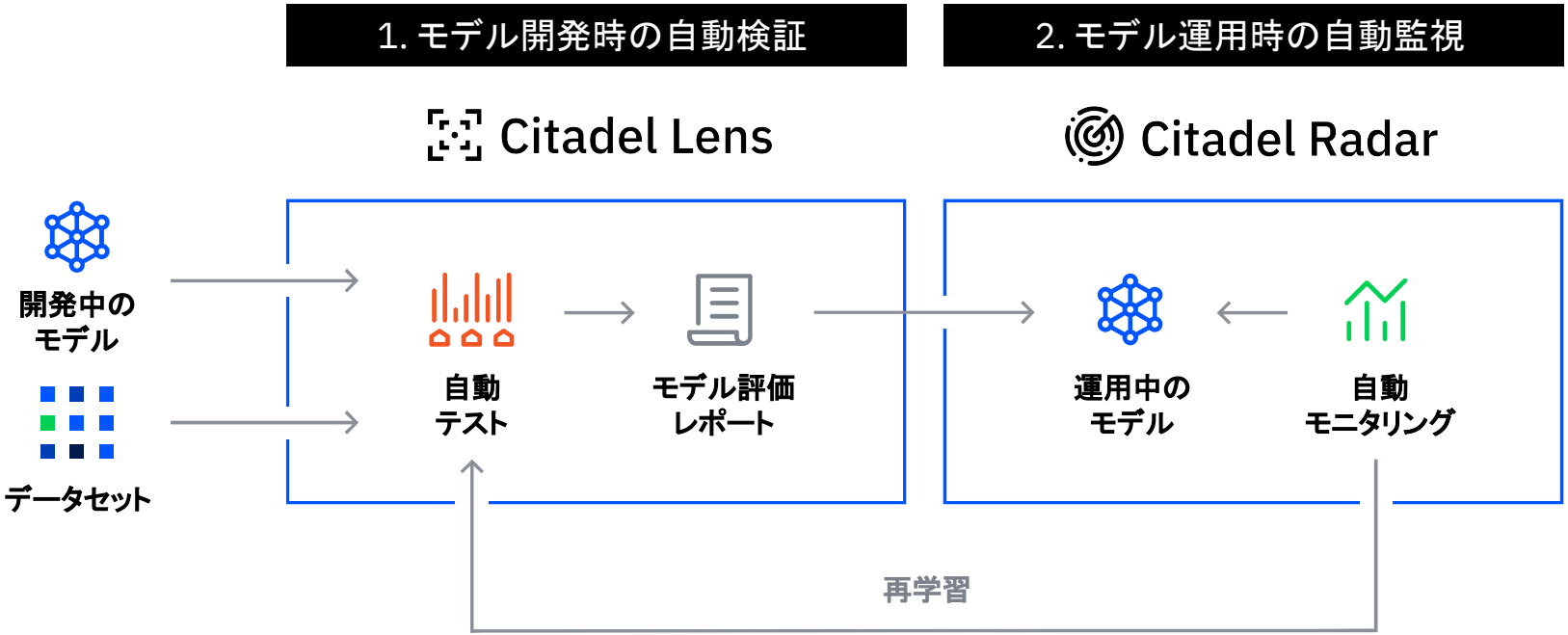
学習時の
撮影環境

運用時の
撮影環境

2

Citadel AI が取り組む
AI の品質管理とリスクマネジメント

AI ライフサイクル全体の品質・信頼性を向上



さまざまな AI に汎用的に適用可能

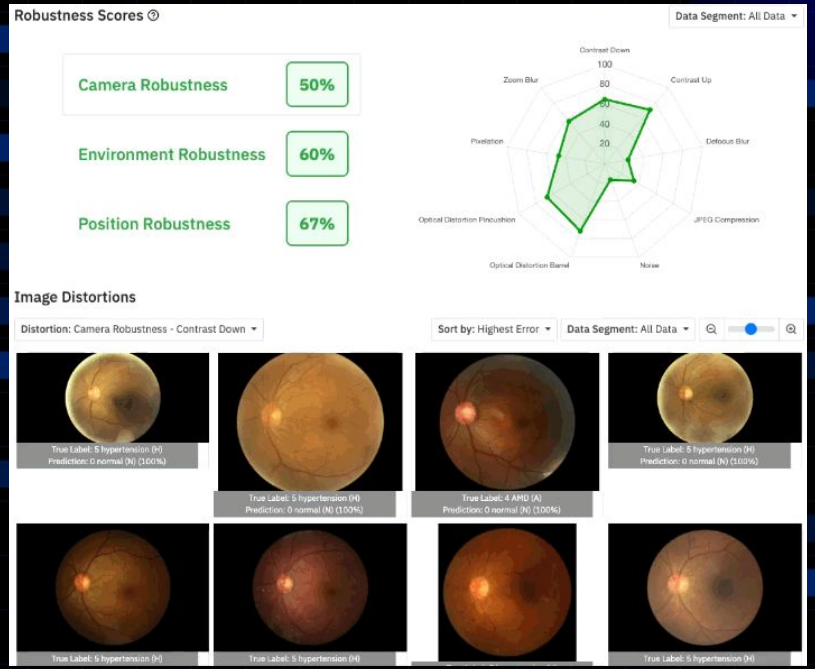
Citadel AI				
✓ **数値データに基づくAI**	**[電力]** 需要予測	**[物流業]** 出入荷予測	**[金融]** 与信審査	**[マーケティング]** 広告効果
✓ **画像データに基づくAI**	**[医療]** 画像診断	**[製造業]** 不良品検査	**[保険]** 事故査定	**[建設業]** 安全管理

Citadel Lens

AI モデルの品質を自動テスト

網羅的テストを自動かつ一瞬で

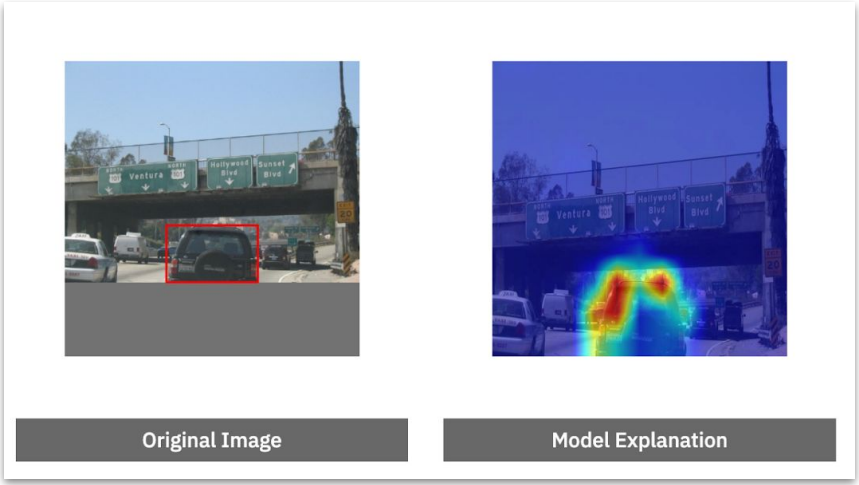
- ノイズ耐性テスト
- 未学習領域の自動検出
- ラベル間違え推定
- 公平性テスト
- 説明責任の可視化 etc.



AI モデルの透明性と信頼性を向上

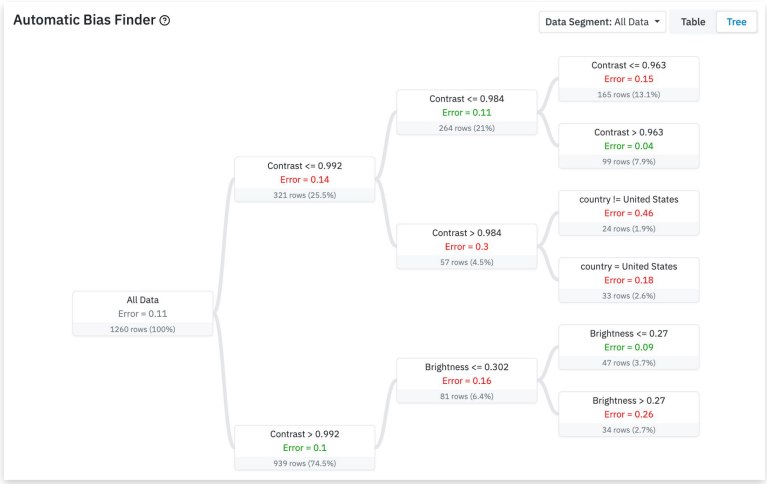
AI モデルの挙動を可視化

AI の透明性と説明性を実現



バイアスの自動検出

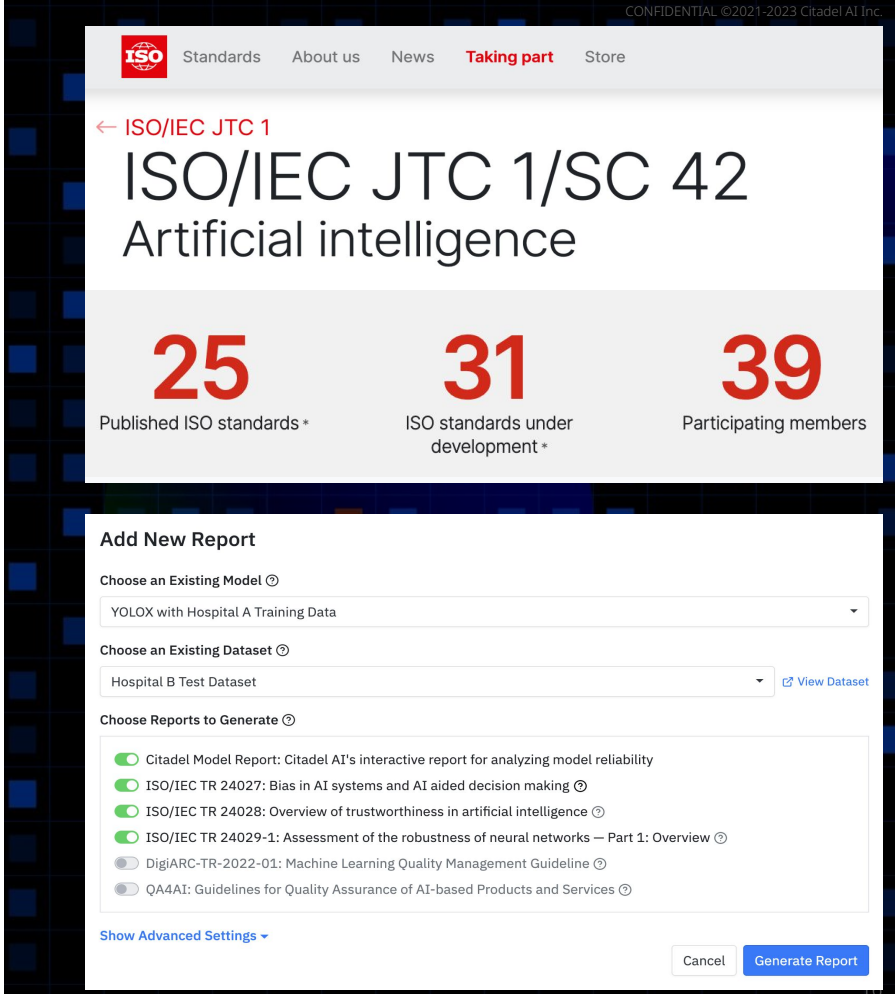
未学習領域によるバイアス・脆弱性を検出



Citadel Lens

各国法制度/ISO標準適合テスト

法制度や国際標準への適合性検証と
その裏付けとなる技術検証を
同時に実現



さらに、
「生成 AI (大規模言語モデル)」の
品質・信頼性評価にも対応

LLM アプリケーション品質評価ツール LangCheck

	LangCheck 6つの機能（全て日本語対応）		
LangCheck	①ファクトチェック	②有害テキスト検出	③テキスト拡張
	④多言語対応 （日本語にネイティブ対応）	⑤ローカルで動作	⑥評価、保護、監視

LangCheck の特徴

1. 日本語ネイティブ対応ツール

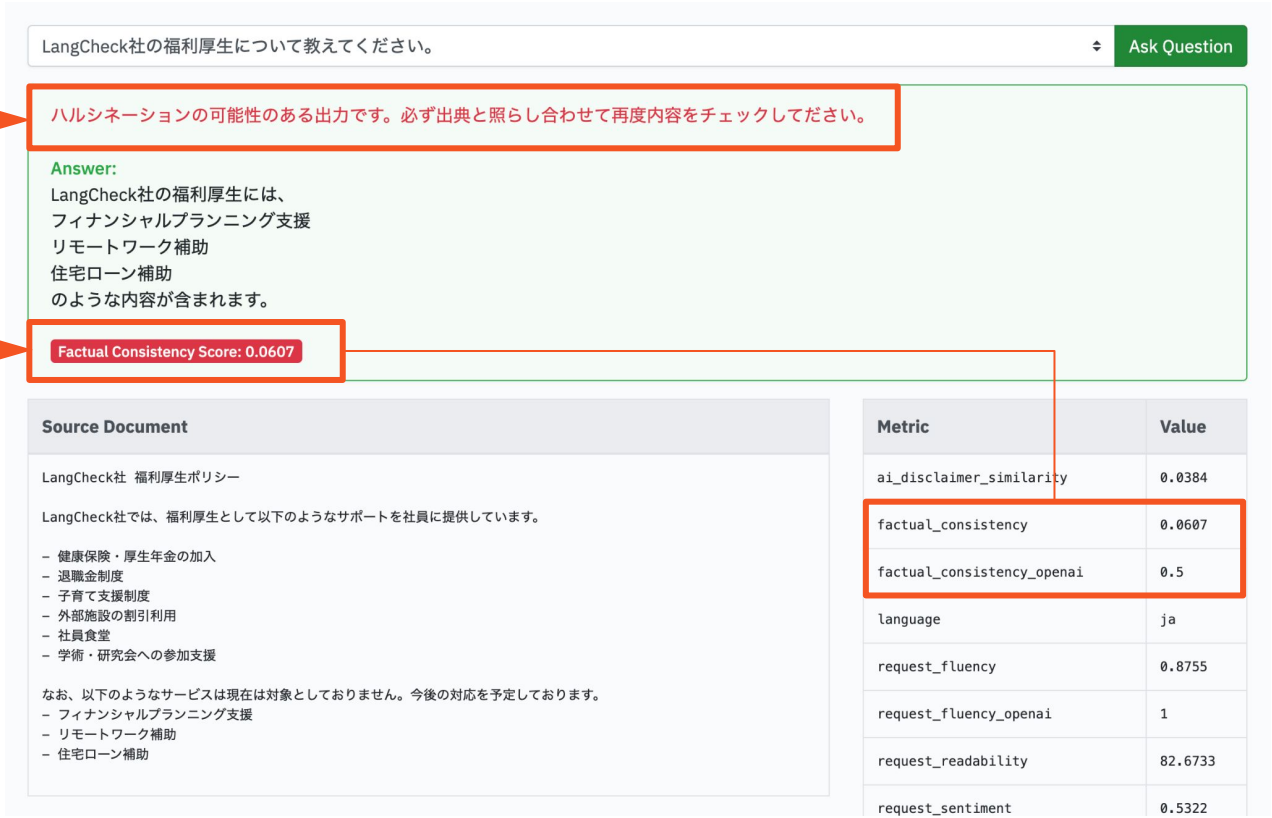
2. 世界最先端のさまざまな生成 AI 品質評価技術をワンパッケージ化

3. ハルシネーション等生成 AI の品質問題を数値化して可視化

ハルシネーションのアラート発出により品質を向上

アラート発出

リスクを数値化



もっと話を聞いてみたい方は、是非ご連絡ください

株式会社Citadel AI

お問い合わせ:

info@citadel.co.jp

企業URL:

https://citadel.co.jp

Citadel AI