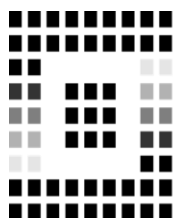


AIセーフティ強化に関する研究開発(株式会社Citadel AI)



Citadel AI

会社名	株式会社Citadel AI
所在地	東京都文京区
設立年	2020年
代表者	小林 裕宜 代表取締役
事業領域・分野	情報・通信
事業年度	2025年度

<事業概要>

本事業では、AIシステムの安全性確保に向け、ガイドラインや評価プロトコル等の共通基盤を整備・公開しました。

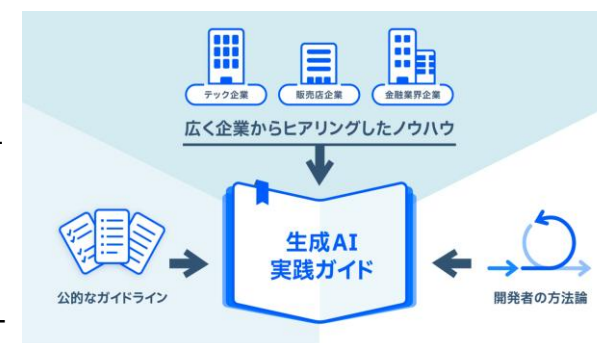
特に弊社では、生成AIやAIエージェントの開発・運用に取り組む企業へのヒアリングを通じ、技術、プロセス、組織文化の観点から共通する実践知を抽出し、「生成AI実践ガイドと企業事例集」として体系化しました。

また、開発者に向けて抽出したノウハウを検証の上で具体化するため、行政手続きなどに関する情報を中心に、外国人などの言語や制度の理解に一定の配慮を要する利用者を想定したチャットボットとして実装し、公開しました。

<得られた成果と活用イメージ>

本成果は、生成AIの利活用とAIセーフティ・AIガバナンスの両立を実現するための方法論をフレームワークとして体系化したものであり、各組織におけるAI導入・運用の参考としての活用を想定しています。

さらに、公開したガイド・事例集とチャットボットの実装を通じて、抽象的な原則を具体的な評価・運用手法として提示し、実務におけるAIセーフティの実践を支援します。



▼本成果に興味をもたれた方はこちらへ

事業者へのお問い合わせ窓口 : info@citadel-ai.com ／事業者Webサイト: <https://citadel-ai.com/ja/>

NEDOへのお問い合わせ窓口: ai-robo_followup@ml.nedo.go.jp