

AI制度研究会 リスク想定と対策について

2024/8/23

株式会社ベネッセホールディングス

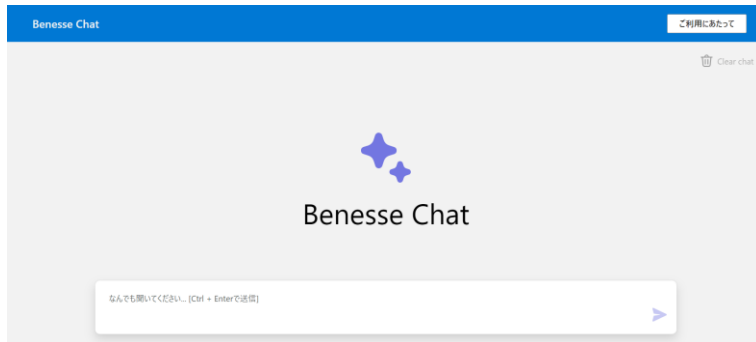
DIGITAL INNOVATION PARTNERS

ベネッセの生成AI活用ステップ

まずはセキュアな実験環境を構築、社内活用から、顧客向けサービス(社外活用)に展開。

Step1

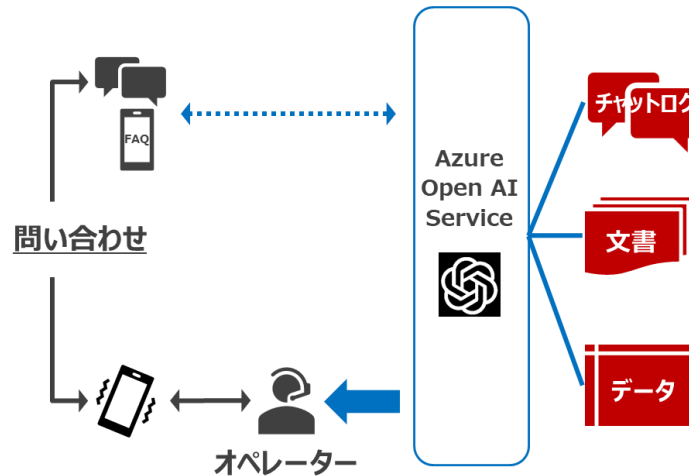
実験環境の構築 (2023年4月～)



グループ社員1万5000人に
BenesseChatを提供し
セキュアな環境で活用を促進

Step2

社内業務の生産性向上 (2023年6月～)



コンタクトセンター業務などにおいて
生成AIを活用し
顧客体験の向上と効率化を実現

Step3

顧客向けサービスの提供 (2023年7月～)



生成AI活用のアイデアを元に
顧客向けサービスを検討
社内チームが開発

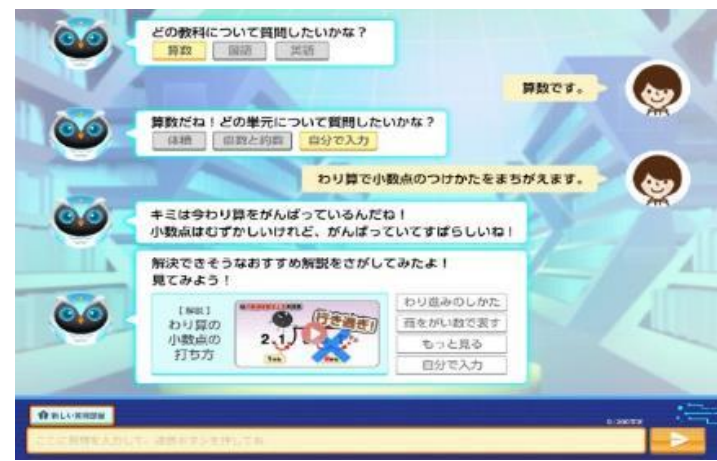
顧客向けサービスの提供

夏休みの「自由研究お助けAI」からはじまり、蓄積したナレッジやデータを活用したサービスを展開中。

自由研究お助けAI (2023年7月)



チャレンジAI質問コーチ (2024年1月)



AIしまじろう (2024年2月)



この日 直近7日間 直近30日間

ゆうくんの気持ち

今日のゆうくんは、とても喜びながらたくさん会話をしていました

うれしー 50%

わくわく 20%

おもしろい 2%

こわい 10%

びっくり！ 10%

いやいや 2%

イライラ 2%

おんんん 2%

会話での興味度

☆☆☆

強い興味を示した会話をみる

興味のある言葉

おさるさん そうさん きりんさん

よく出てきた言葉

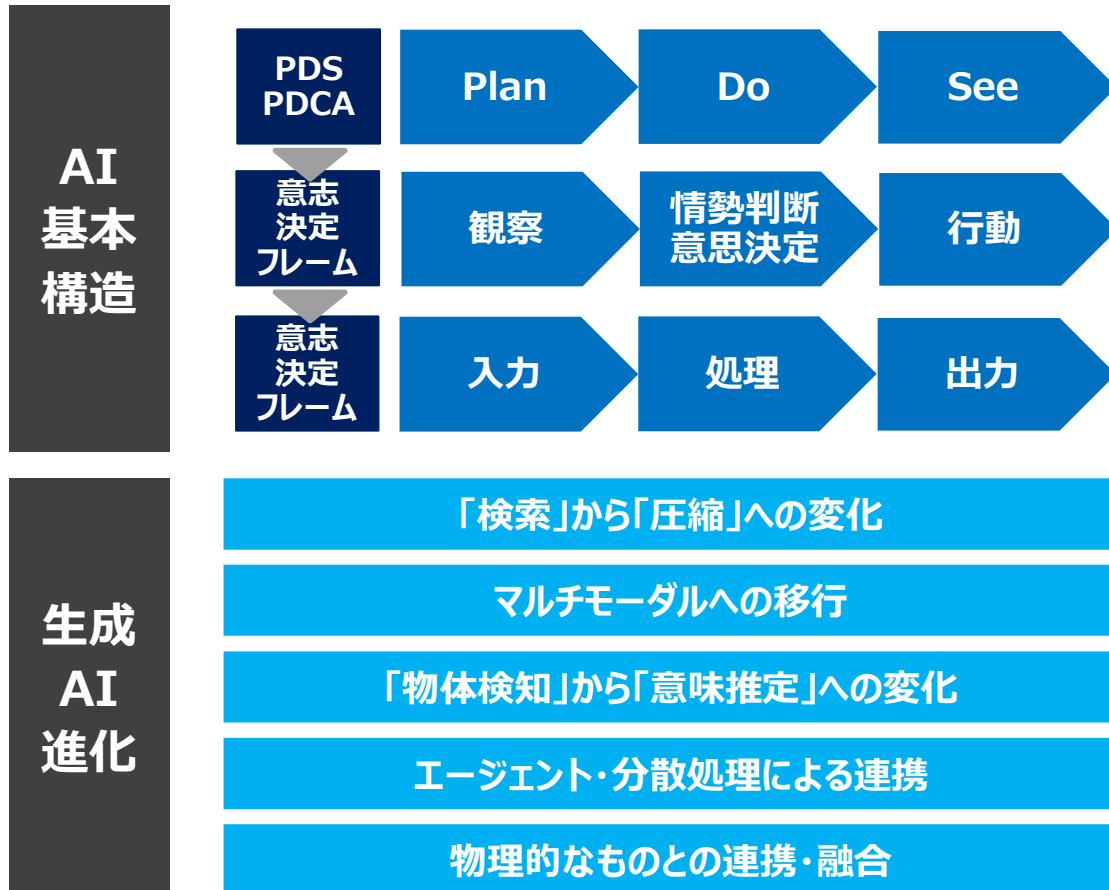
1 恐竜	18回
2 幼稚園	17回
3 ママ	16回
4 ご飯	15回

想定されるリスクについて

AI利用者の視点から見た、「技術」と「リスク」の見立て。進化が加速するなかで、前提変化も踏まえ、向き合う必要がある。

技術の見立て

基本構造を持ちながら、さらに進化が加速する見立て



リスクの見立て

技術進化のなかで、リスクポイントも拡大する見立て

■ 基本構造からみるリスクポイント

- ・「入力」「出力」での、「著作権」「情報セキュリティ」への配慮
- ・グローバルの地域ごと、利用者年齢ごとの配慮

■ 生成AIの進化に伴うリスクの拡大

- [圧縮] 利用データのブラックボックス化や偏り、ハルシネーション
- [マルチモーダル] 画像・動画・音声などリスク対象の拡大
- [意味推定] バイアスによる疎外・差別の誘発
- [エージェント・分散処理] 責任主体の分散・不明瞭化
- [物理連携・融合] 物理的な危害への拡大

サービス価値向上を目的に、社会実装での具体策が重要

リスクに対する対応について

基本構造からリスクポイントを明確化し、リスク対策を実装。技術進化による変化を前提とした、リスク対策が重要。



■ AI活用は事業活動に埋め込まれていく

⇒ 事業特性に応じた既存法対応

■ 生成AI技術は進化が加速する

⇒ 技術・性能評価の基準・デファクトの整備

■ グローバルなサービス展開が必要

⇒ 国際整合性と展開しやすい地域の拡大

何をするか

自由研究テーマ
を考えて提案する

自由研究の理科系テーマ
でおすすめは？

処理の工夫（＝優位性）

「虫の観察」等、身近に観
察できるテーマを提案

出力されるものが、顧客の期待値を超える、役立つものが
出力されるものに、リスクがないか

何をしないか

答えを教えない
変な質問に答えない

読書感想文を書いて

読書感想文は書けないよ

入力

処理

出力