

AIエージェントシステムを用いた組織的な学習 への取り組み

8th Grand Canvas : AI品質の未来を共に描く
～AI品質マネジメントネットワーキングシンポジウム～

杉山 阿聖 (Citadel AI)

2026/7/16

目次

1. はじめに
2. 背景と問題意識
3. 先行事例
4. 研究方法
5. 先行知見の整理と中心仮説

6. 評価設計
 7. 現時点の検証結果
 8. 今後の取り組み
 9. 結論
- 問い合わせ先

1. はじめに

- 生成AIの急速な普及 → コーディング支援・文書作成では実践例が蓄積
- 一方、組織の意思決定や企画プロセス(プロダクトディスカバリー)への AI活用は研究・実践例が少ない
- 本研究: Citadel AIにおける企画プロセスへの AIエージェント活用を探索的に実践
- リーンスタートアップ型の仮説検証サイクルを研究プロセスに適用

仮説を立て → 最小限の実験 → 結果を観察 → 示唆を整理 → 次の仮説を更新

2. 背景: **AI**の生産性パラドックス

- AIEージェントの組織的活用ノウハウは未だ十分に体系化されていない
- 品質保証・リスク管理の手法が未確立 → PoCから本番運用への「壁」
- 個人レベルの効率化 ≠ 組織・事業レベルの成果への転換
- AIへの期待と実際の事業成果のギャップ = 「AIの生産性パラドックス」

個人の効率化効果は報告される一方で、
それが組織・事業レベルの成果に転換されにくい構造的課題

2. 背景: **DeNA**の事例

DeNA「AIオールイン」宣言(2025年)

- 約3,000名の現業部門でAI効率化を推進、半分の人員で現業を維持 → 残り人員を新規事業へ
- 成果: 開発エンジニア生産性 最大20倍、法務リーガルチェック工数 90%削減
- しかし「タスク工数が減ると余った時間で新たな仕事を追加してしまう」(南場 , 2026)

示唆: AIが「設計・実装・運用」の効率を高めても、
「企画」プロセス(顧客理解・仮説形成)は自動的に前進しない
→ 企画プロセスそのものに AIを組み込む必要がある

2. 背景:グローバルにみた共通課題

- McKinsey "The State of AI 2025": AI活用は広がるが依然パイロット段階に留まる
- 企画の「上流」(顧客理解・市場定義・価値仮説形成)での AI活用報告は少ない
- SaaStr Annual 2026:「Agent Parity(エージェント実装精度の横並び化)」の見通し
- AI機能の均質化が進む前に、実装の質と深度を高めることが組織の競争力を左右する

「今後12ヶ月で最高のエージェントを作れた者が勝つ」— Jason Lemkin (SaaStr CEO)

2. 背景: **Citadel AI**を取り巻く事業環境

- 当初: 機械学習のテスト・モニタリングに関するプロダクトを販売
- 生成AIの登場 → AIセーフティ・AIガバナンスへのニーズが高まり、注力分野をシフト
- 課題: どの顧客のどの課題に、どのような価値を提供すべきかを改めて明確にする必要
- 顧客ペルソナについて組織内で一致した見解を持てていない
- 顧客が持つ課題についても統一された見解が不十分

3. 先行事例: **DevOps** → **MLOps** → **GenAIOps**

DevOps

Dev+Opsの協調
多能工化

MLOps

データサイエンティスト
+MLエンジニア追加

GenAIOps

業務ドメイン専門家を
チームに組み込む

GenAIOpsの特徴

- 生成AIは「対象業務」「技術」「提供価値」の三層の不確実性を持つ
- ドメイン専門家との直接的な協働なしに品質・安全性の向上は難しい

3. 先行事例:コーディングエージェントの高度化

- GitHub Copilot: タスク完了速度 55.8% 向上 (Peng et al., 2023)
- Devin (Cognition Labs, 2024): 自律的にコード生成・テスト・PR提出
- SWE-bench解決率: 13.86%(2024/3) → 約75%(2025年最高水準)
- 「設計・実装」領域での AIエージェント活用は急速に進展

**対比: 上流の企画プロセス(顧客理解・仮説形成・価値判断)への
AI活用は依然として空白領域**

3. 先行事例: 企画領域での**AI**活用の空白



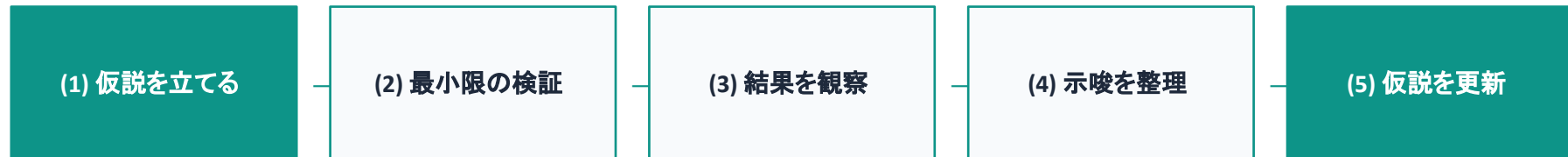
- 顧客理解・価値仮説の形成・事業ドメイン知識の統合は非定型的・探索的
- 自動化や定式化になじみにくいとされてきた
- この「企画プロセスへの AI活用の空白」が本研究の取り組む問い

3. 先行事例: 業務ドメイン知識の統合が鍵

- GenAIOps調査: AIエージェントの提供価値を高めるには業務ドメイン知識の統合が決定的に重要
- 設計・実装領域のみならず、あらゆる業務領域に共通する示唆

「業務の価値をもっとも理解している専門家と開発チームが中心となって開発を進めること自体が、
もっとも確実で本質的なリスクへの対応となる」
—— 杉山ら(2025)

5. 研究方法:リーンスタートアップ型の探索的実践研究



研究テーマの変遷

- 初期: 顧客AI・PO AIの開発
- 初回検証: 顧客AIとPO AIの対話実験
- 更新: AIエージェント同士の会話による企画プロセスの観察
- 現在: Slackワイガヤ実験の検証設計

5. 先行知見: 業務専門家協働モデルの示唆

- GenAIOps調査(20社超のヒアリング): 業務ドメイン知識の統合が決定的に重要
- 業務専門家を開発チームに組み込み、プロンプト記述・評価データ構築に参加させるチームが高い成果
- 本研究の動機と強く呼応: 事業ドメイン知識 × 役割定義の組み合わせが提案品質を高める
- 「ドメイン専門家との協働が提供価値を高める」という実践的知見と理論的に整合

5. 先行知見: エージェント均質化後の差別化要因

- SaaStr Annual 2026: 差別化はエージェント自体ではなく、与えるドメイン知識・顧客文脈・人間の介在にある
- 「差別化は業務領域への入り込み方、データの蓄積、顧客現場の FDE (Forward Deployed Engineer) にある」
- Monaco (AI-Native CRM): 経験豊富な人材が顧客側に直接張り付く「人間ガイド型」モデル
- エージェントの品質はそこに与えるデータの質を超えることはない
- Knowledge Base・セマンティックレイヤーの整備がエージェント活用の前提条件

5. 中心仮説

AIエージェントを活用したワイガヤ
—— 複数の役割を持つエージェント同士の対話 ——
これを企画プロセスに導入することで、
組織としての顧客理解・仮説形成・機能提案・価値判断の
質と速度を高めることができる。

ワイガヤとは

立場や役職にとらわれず、まだ答えが決まっていないテーマについて自由に意見を出し合い、対話を通じて本質や方向性を探る場。AIエージェント同士の会話で事前シミュレーションし、人間が観察することでより良い問いや仮説を見つけられるのではないかな。

6. 評価設計

Stage 1: 顧客AIの実現可能性

評価者: CxO・FDE
(顧客理解・事業ドメインの専門家)

AIエージェントの対話を手動で操作している所を見て評価してもらう



Stage 2: マルチエージェントのユーザーテスト (Slackワイガヤ実験)

評価者: 英語話者エンジニア
(1) 事前アンケート → (2) 会話観察
→ (3) 自由質問 → (4) 事後アンケート
→ (5) 全体振り返り

評価観点

- 顧客理解が深まったか
- 顧客課題をより具体的に説明できるようになったか
- プロダクトに実装すべき機能の改善点が見つかったか
- 継続して使いたいと思うか
- CxO・FDEが見て、機能提案や仮説の筋が良いと評価できるか

7. 結果: Stage1.顧客AIの実現可能性

エージェント	役割	与えた情報
顧客AI	顧客らしい反応を返す	実際のヒアリング記録、顧客理解情報
PO-A	POとして顧客と会話し課題・方向性を考える	POとしての役割、会社のビジョン
PO-B	PO-Aと同じ役割	PO-A + AIガバナンスの知識(事業ドメイン)
ファシリテータ (試験運用中 → 中止)	各エージェントによる議論結果の総評(論点抽出、 懸念事項共有、NextAction提案) (各エージェントに対する会話振り)	ファシリテータとしての役割

- 各エージェントの出力の妥当性や質を Citadel AIの社員の主観で評価
- 顧客AIにおいては出力に対しての想定以上の妥当感、納得感があると評価
- 各POも双方妥当感の問題ないレベルであったが、提案の質は Bの方が高く見られた
- ファシリテータは、試験的に運用。総評の質は高評価であったが、議論誘導の方向性に懸念が生じる結果が見られたため、Slackワイガヤ(stage2)での検証は中止。
- 総評に対しては参照者が「ワイガヤの経緯を追い、示唆を見つけ、認知」というプロセスを省略できる点は有用と評価

7. 結果:初回検証の位置づけ(**Stage1**)

- 中心仮説(複数エージェントのワイガヤ)の直接検証ではなく、構成要素の確認
- 顧客AIとPO AIの1対1対話の有効性を確認する前段階実験
- この結果がSlackワイガヤ実験設計への方向転換の根拠

Slackワイガヤ実験(Stage2)の結果はp.31にて後述

7. 結果: Stage1 - 1

①: 現場に近い会話をAI上で再現できる可能性

- PO-Aと顧客AIの会話は、Citadel AI社内での普通の会話にかなり近いものとして再現
- 現時点で存在する役割の人が持っている情報を各AIエージェントに与えることで、現場の社内会話に近い状況を再現可能

単なる資料要約だけでなく、自分たちの会話や思考の癖を外側から観察できるという示唆

②: 事業ドメイン知識が提案品質に影響

PO-A

POの役割 + 会社のビジョン
→ 会話はリアルで人間らしい
→ 提案の具体性は限定的

PO-B

PO-A + AIガバナンスの知識
→ より論理的で詳細な情報を提供
→ 提案の質が明らかに向上

- PO AIの質は役割定義と会社ビジョンだけでは不十分。事業ドメイン知識を組み合わせることで初めて提案品質が高まる。
- 「チーム自身が顧客・事業・業界を深く理解することが重要」という実践知は、AIエージェントにも当てはまる。

7. 結果: Stage1 - 2

③: AI同士の会話と人間同士の対話のギャップ

- PO-A: 会話がリアルで人間らしい
- PO-B: より論理的で詳細だが、人間にとって受け取りやすいとは限らない
- AIエージェント同士の会話で有効なやり取り ≠ 人間同士の対話として再現可能
- 人間同士の場では情報量・受け取り手の疲れ・関係性・モチベーションが影響

現時点での位置づけ: AIエージェント同士の会話は人間の会話の代替ではなく、論点や仮説を事前に可視化し、人間が検討する材料を増やすための補助的な仕組みである

④: 思考・対話の観察装置としての可能性

AIエージェント同士の会話を観察することで以下を外側から見る事が可能:

- どこで論点が噛み合っていないか
- 顧客理解がどこで浅くなるか
- プロダクト・サービスの方向性がどこで曖昧になるか
- どの情報が足りないと提案の質が下がるか
- どの情報を追加すると議論の質が上がるか

論点の整合、顧客の理解度、提案の質などの価値判断基準をAIエージェントからユーザ自身が獲得できる。その基準をもって、ユーザが AI提案の品質を一定検証可能となることで、対話の質も上がり、学習の好循環が生まれる可能性がある

7. 結果: Stage2. マルチエージェントのユーザーテスト

事前→事後の評価結果				
Form_Responses				
Timestamp	Email Address	Use of Test Results / テスト結果	Pre-Test Questionnaire / 事前ア	Post-Test Questionnaire / 事後ア
6/24/2026 12:15:25		Yes / はい	2	3
6/24/2026 19:02:43		Yes / はい	2	3
6/25/2026 14:55:26		Yes / はい	1	5
6/29/2026 16:41:17		Yes / はい	1	5
6/30/2026 11:56:30		Yes / はい	2	3

非日本語話者

- 総論として、一定の理解度向上効果が見られた
- 特に、非日本語話者における理解度向上が顕著に表出
- 日本の商習慣・ビジネスカルチャー等の理解獲得が大きいと分析
- 例: 判断の持ち帰り、キーパーソンの不参加、曖昧な要望、クライアントの課題感・人となり・価値観
- 設問: 「現時点で、今回のテーマについて、顧客の本音・本質的な課題・次に検証すべき論点を具体的に説明できる状態だと思いますか？」
- 1=ほとんど説明できない 5=かなり具体的に説明できる

8. 今後の取り組み

社内仮説検証・AIセーフティに係る具体ユースケースの考察

- 一定のSlackワイガヤ継続
- エージェント選択の使いやすさ
- 会話ログの観察・振り返り活用
- 人間の途中介入の有効性
- アンケート・結果分析
- 仮説の確からしさの検証
- テストランに向けたケーススタディ・検証項目の考察

企業でのテストラン及びAIセーフティネタの発掘

- 本取組の紹介
- 導入・運用支援
- 実践・試験運用
- テストラン結果・アンケート分析

フレームワーク整理と対外発表

- 実践的フレームワーク化
- 論文・対外発表
- 外部企業での検証
- 知見の一般化可能性の向上

9. 結論(暫定)

検証で得られた3つの示唆

1

顧客情報と役割情報を適切に与えることで、
現場に近い会話をAI上で再現できる可能性

2

事業ドメイン知識の付与が
PO AIの提案品質を向上させる可能性

3

AIエージェント同士の会話が
組織の思考・対話を観察する装置として機能し得る可能性

今後: Slackワイガヤ実験の実装と評価を通じて仮説を実証的に検証

参考文献

Dohmke, T., Iansiti, M., & Richards, J. (2023). Sea Change in Software Development. arXiv:2306.15033.

McKinsey & Company (2025). The State of AI 2025.

Peng, S. et al. (2023). The Impact of AI on Developer Productivity. arXiv:2302.06590.

Yang, J. et al. (2024). SWE-agent. NeurIPS 2024.

SWE-bench leaderboard (2025). swebench.com

杉山阿聖ほか(2025). 生成AI実践ガイドと企業事例集 NEDO委託業務成果物

野中郁次郎・竹内弘高(1996). 知識創造企業 東洋経済新報社

工藤真由(2026). 米国AI駆動GTMの最前線ー SaaS AI 2026.

南場智子(2026). DeNA AI Day 2026 クロージングセッション.

問い合わせ先

研究対象

- Citadel AI社内における企画プロセスへの AIエージェント活用
- 顧客AIとPO AI(プロダクトオーナー)を用いたマルチエージェント対話の設計・実施・観察

杉山 阿聖

Mail: asei@citadel-ai.com

AI技術・AIガバナンスの文脈 × リーンスタートアップ・アジャイル・組織学習の文脈