

ソフトウェア品質シンポジウム2024 受賞研究の紹介

ソフトウェアレビューにおける生成AI活用の研究

～ChatGPTが欠陥検出と指摘伝達をアシスト～

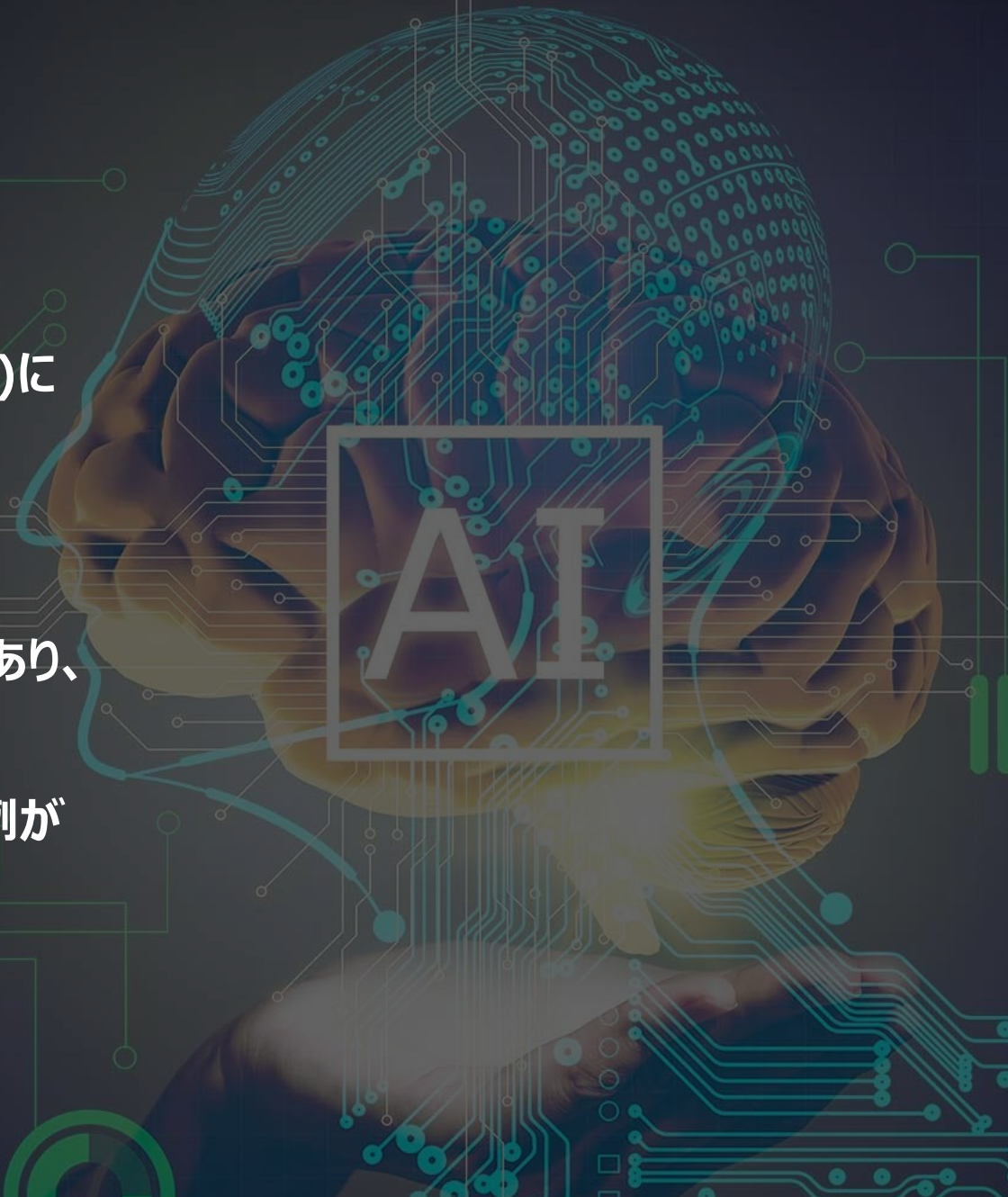
2023年度 ソフトウェア品質管理研究会
研究コースⅡ ソフトウェアレビュー

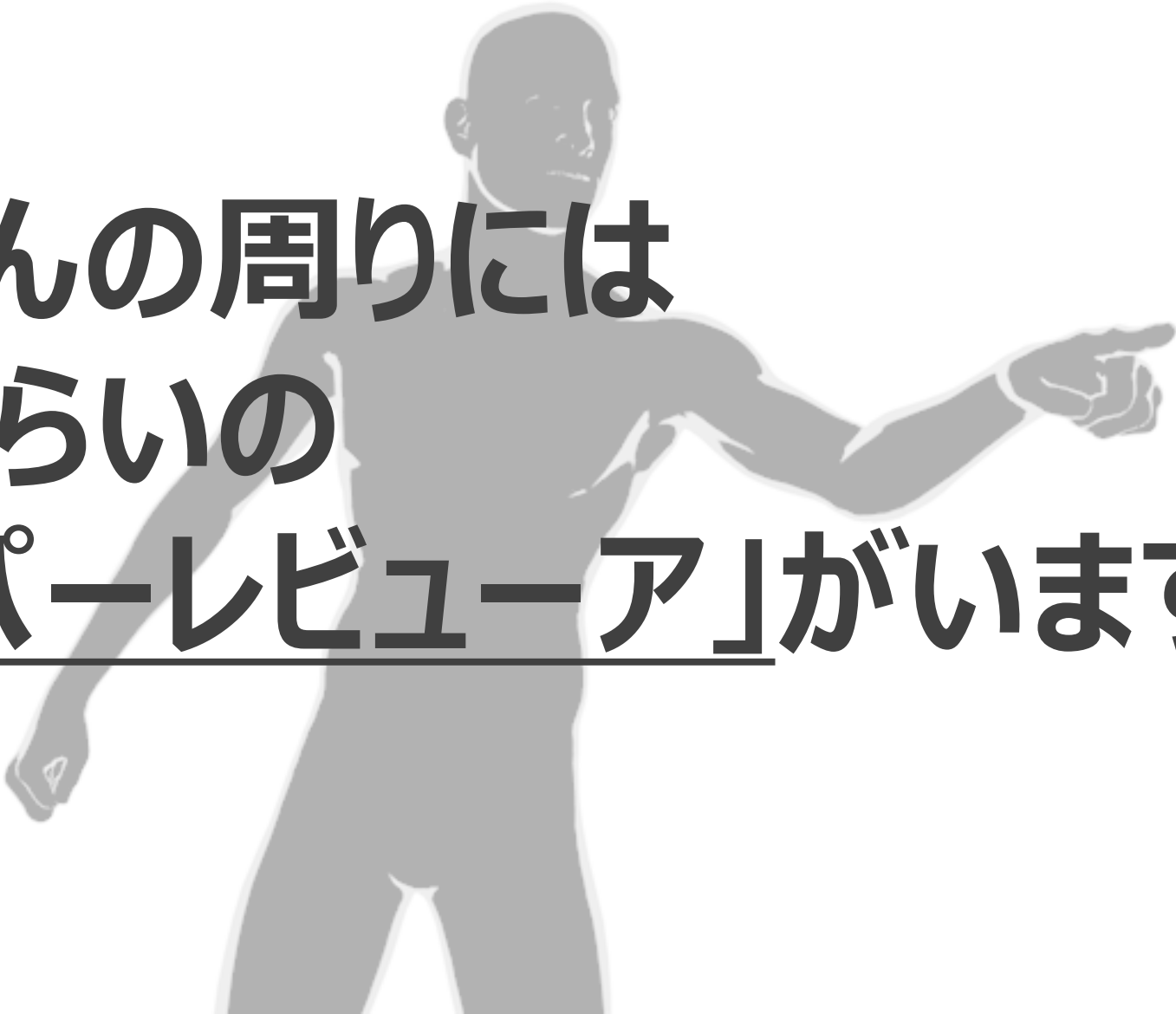
研究員 : 北里 竜 (ブライシス株式会社)
片桐 汐駿 (アズビル株式会社)
馬場 大輔 (株式会社オージス総研)
星野 智彦 (株式会社アイシン)

主査 : 中谷 一樹 (TIS株式会社)
副主査 : 上田 裕之 (株式会社DTSインサイト)
アドバイザー : 安達 賢二 (株式会社HBA)

本研究に至った背景

- ・ 2023年度 SQiP研究会 研究コースⅡ(ソフトウェアレビュー)に参加した研究員が抱える共通の課題として、レビューの質がレビューアのス킬に依存するという「レビューの属人化」の問題があった。
- ・ 研究開始時、ChatGPTをはじめとする生成AIがトレンドであり、ChatGPTを研究に絡めたい思いがあった。
- ・ レビューの分野では、生成AIをどのように活用すべきかの前例がまだ少ないため、取り組む必要性を感じた。





みなさんの周りには
何人ぐらいの
「スーパーレビューア」がいますか？



スーパーレビューア

- 重大欠陥もビシバシ見つける
 - 的確な指摘伝達
-

普通のレビューア

- 軽微な欠陥ばかり...
 - 指摘もちゃんと直してくれない...
-





スーパーレビューアに育つまでは
途方もない時間と労力がかかる

SQIP研究会の既存研究



- 重大欠陥もビシバシ見つける
- 的確な指摘伝達

レビューへの効果が実証された
過去の手法のほんの一部...

欠陥検出

WUT法

レビューア向け思考能力トレーニング法



レビューアに必要な
「要約力」と「仮説力」
をトレーニングで向上

指摘伝達

RCS法

レビューコミュニケーションスタイル法



受け手のコミュニケーション
スタイルに合った
指摘伝達

要約力、仮説力が
必要！

想像力・言語能力
が必要

トレーニングが必
要！

ハードルだらけ
じゃねーか！



ここで救世主 ChatGPTの 登場！！！！



ChatGPTの強み



①言語能力

②応答速度



③幅広い知識

まさに百人力！！！！

ChatGPTの弱み



でも時には
嘘をつくことも・・・



適切な使い方ってなんだろう



代わりに欠陥を検出する

代わりに指摘を伝える

全て任せるのは怖いかも・・・



ChatGPTが全力アシストします！

生成AI活用のイメージ

パワーアシストスーツ Power assist suit

スーパーレビューアと同じことができるように
レビューアを**アシスト**します。



提案

SQIP研究会の既存研究



- 重大欠陥もビシバシ見つける
- 的確な指摘伝達

提案手法は2つの既存研究を
ChatGPTにアシストしてもらう手
法です

欠陥検出

WUT法

レビューア向け思考能力トレーニング法



レビューアに必要な
「要約力」と「仮説力」
をトレーニングで向上

指摘伝達

RCS法

レビューコミュニケーションスタイル法



受け手のコミュニケーション
スタイルに合った
指摘伝達

欠陥検出手法の説明

先行研究であるWUT 法では、以下の手順で欠陥を検出する。

1. レビュー対象の特徴を掴む（課題：要約力が必要）
2. レビュー対象に混入されている欠陥について仮説を立てる（課題：仮説力が必要）
3. 実際にレビューを実施して指摘する

この手順のうち、1.と2.を生成AI にアシストさせる。

要約アシストの具体例

プロンプト

下記の要求仕様書について,初めて見る人が内容を把握できるように,背景も含め,箇条書きで要約をお願いします。



ChatGPT
3.5



【要約】 アンケート集計業務改善システム

****A社アンケート集計業務改善システム 開発要求仕様書 要約****

- A社では社内事務処理コスト削減のため,アンケート集計業務のシステム化が必要。
- アンケート実施は月4回で,手動の集計によるコストと時間の浪費が課題。
- システム化により業務効率向上と集計結果の早期取得を目指す。

1. **社員意識調査**

- 月1回、全社員300人
- 電子メールでエクセルシートを送付し、回答をメールで返信。
- 人事部でエクセルへ転記し手動集計。

2. **社内研修参加者アンケート**

- 月3回、各回20人
- 紙のアンケート用紙に記入後、回収・手渡し。
- 人事部でエクセルへデータ入力し手動集計。

1. 概要

- 社内WEB環境にアンケート記入システムを構築。
- 管理者と社員用の機能を備え,他システムとの連携については詳細未記載。

2. 機能説明

2.1 ログイン機能

- ログインは社員IDと専用パスワードで行う。
- パスワードは随時変更可能。

2.2 管理者用機能

- **アンケート一覧表示**

- 登録済みアンケートを一覧表示。

- **アンケート登録**

- タイトル、種類、設問数、設問内容を入力し、作成。
- 種類により自動で参加者情報を取得。

- **アンケート集計**

- ボタン押下でアンケートが集計され、評価別件数の表と棒グラフ表示。

- **アンケート詳細表示**

- 各設問の回答詳細を表示。

2.3 社員用機能

- **アンケート一覧**

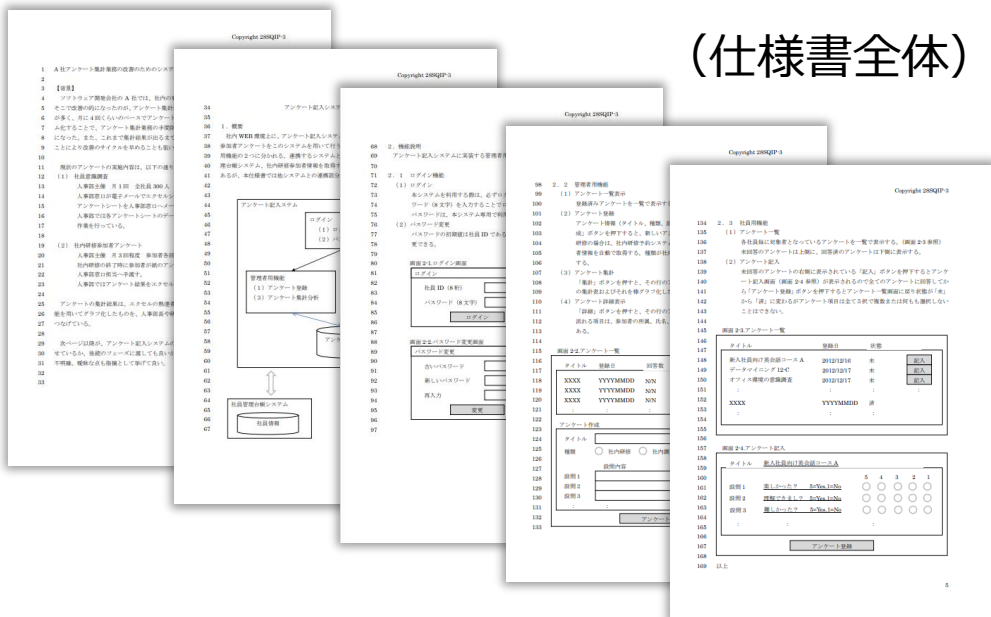
- 未回答の上部と回答済の下部に分類して表示。

- **アンケート記入**

- 未回答のアンケートに回答し、登録ボタンで状態を変更。
- 複数選択必須。

アンケート集計業務改善システムの要求仕様書

(仕様書全体)



仮説アシストの具体例

プロンプト

あなたは優秀なシステムエンジニアである。今回、下記のシステムの開発を行うことになり、システム化の背景をまとめた。このシステムに必要な画面と、それぞれの画面の仕様を挙げてほしい。



ChatGPT
3.5



アンケート集計業務改善システムの要求仕様書

(背景のみ)

Copyright 2025 QIP-3

1 A社アンケート集計業務の改善のためのシステム開発
2
3 【背景】
4 システム開発開始前のA社では、社内の事務処理コストの削減を行うことになった。
5 そこで効果的になったのは、アンケート集計業務である。A社ではアンケートを取るこ
6 が多く、月に4回くらいのペースでアンケートを実施している。その業務の一部をシス
7 テム化することで、アンケート集計業務の手間削減および事務処理コストの削減を図ること
8 になった。また、これまで集計結果が出るまでに2週間近くかかっていたのを迅速化する
9 ことにより業務のサイクルを早めることも狙いとしている。
10
11 現状のアンケートの実施内容は、以下の通りである。
12 (1) 社員意識調査
13 人事課主催 月1回 全社員 500人
14 人事課が電子メールでエクセルシートを全社員に送付し、各社員が記入後
15 アンケートシートを人事課窓口へメールで返送する。
16 人事課では各アンケートシートのデータを1つのエクセルシートに転記して集計
17 作業を行っている。
18
19 (2) 社内研修参加者アンケート
20 人事課主催 月3回程度 参加者各回 20人程度
21 社内研修の終了時に参加者が紙のアンケート用紙に記入し、講師が回収して
22 人事課窓口担当へ手渡す。
23 人事課ではアンケート結果をエクセルにデータ入力して集計作業を行っている。
24
25 アンケートの集計結果は、エクセルの表形式である人事課窓口担当がエクセルの標準機
26 能を用いてグラフ化したものを、人事課長や研修講師が研修や傾向を見て分析を行い改善に
27 つなげている。
28
29 表ページ以降は、アンケート記入システムの要求仕様書である。ビジネスユースを課
30 定しているが、後続のフェーズに属しても良いという観点でレビューを付けて頂きたい。
31 不明点、疑問なども指摘として挙げて良い。
32
33

1

【仮説】 アンケート集計業務改善システム

システムの開発に必要な画面および仕様は以下の通りです：

1. **アンケート作成画面**

- 社員意識調査と社内研修参加者アンケートの作成が可能
- 質問の追加、変更、削除ができる

2. **アンケート送付画面**

- 社員意識調査：全社員に対して電子メールでアンケートを送付
- 社内研修参加者アンケート：参加者に対して終了後に電子メールでアンケートを送付

3. **アンケート回答受信画面**

- 社員意識調査：社員がアンケートを記入し、回答を電子メールで送信
- 社内研修参加者アンケート：紙のアンケート用紙からデータを手動で入力

4. **集計結果表示画面**

- アンケートごとの集計結果を表示
- グラフ化した推移や傾向を含む詳細な分析が可能

5. **通知お知らせ画面**

- 集計が完了した際に関係者に通知
- グラフや分析結果が利用可能である旨を通知

以上が、A社アンケート集計業務改善のためのシステムの主要な画面と仕様です。

これによりアンケートの作成、送付、回答受信、集計、結果表示までの一連のプロセスが効率的に管理できるでしょう。

指摘伝達の手法の説明

先行研究であるRCS 法では、以下の手順で指摘を伝達する。

1. 作成者のコミュニケーションスタイルを判別する
2. 欠陥を検出し、指摘事項を記載する
3. 作成者のコミュニケーションスタイルに適した表現で指摘を伝える

（課題：想像力・言語能力が必要）

この手順のうち、3.を生成AI にアシストさせる。

プロンプト

人のコミュニケーションスタイルにはいくつかのパターンがあります。まず、決断が速いか遅いかの2パターンに分けられます。さらに、理論派か感情派かの2パターンに分けられます。つまり、全4パターンに分けられます。次のコメントを、[RCSの種類]の人に伝わりやすい言い方で言い換えてください。

RCSの種類

デシジョンRCS
クリエイティブRCS
ロジカルRCS
エモーションRCS

コメント（言い換え前）

「炒飯が人気のある料理として注目されているユーザー」というのは、「炒飯が好きな人」なのか「炒飯が看板商品である店」なのか**わからない**。」



ChatGPT
3.5

言い換え後の、「どちらかを教えていただけると、より具体的に理解できるので助かります。」という言い回しが、へりくだった感じで好感をもてる。
言い換え後は、正論を振りかざしていない感じ(正解は誰にもわからないので、そういう態度のほうが好感をもてる)

コメント（言い換え後）

「炒飯が人気のある料理として注目されているユーザー」というのは、「炒飯が好きな人」なのか「炒飯が看板商品である店」なのか、**私にはよくわかりません。**
どちらかを教えていただけると、より具体的に理解できるので助かります。

検証

検証

ChatGPTのアシストであなとも
スーパーレビューアになれる

①欠陥検出

②指摘伝達

実験

既存のレビュー手法を
ChatGPTにアシストさせてみた

検証

ChatGPTのアシストであなとも
スーパーレビューアになれる

①欠陥検出

②指摘伝達

欠陥検出手法の説明

先行研究であるWUT 法では、以下の手順で欠陥を検出する。

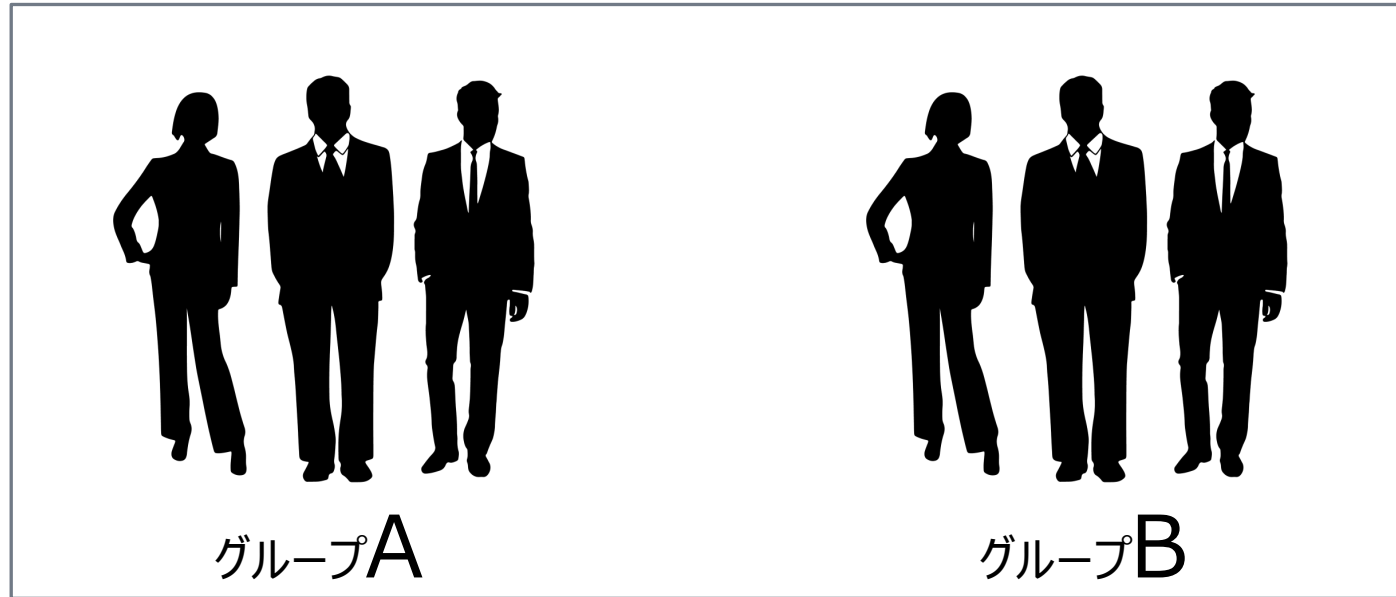
- 1.レビュー対象の特徴を掴む（課題：要約力が必要）
- 2.レビュー対象に混入されている欠陥について仮説を立てる（課題：仮説力が必要）
- 3.実際にレビューを実施して指摘する

この手順のうち、1と2を生成AI にアシストさせる方法を提案する。

実験目的

ChatGPTのアシストで、
重大な欠陥の検出数を増やす

実験内容



研究員の所属企業から6名選出し、
3名ずつAグループ/Bグループに分ける

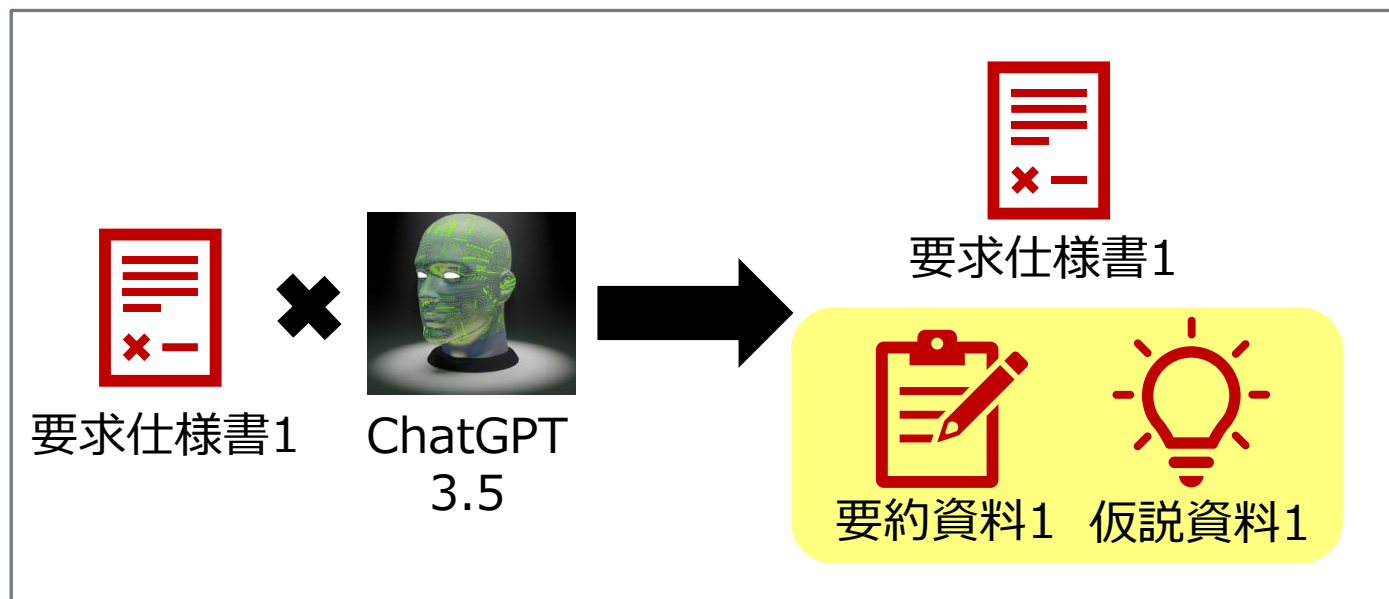


要求仕様書1



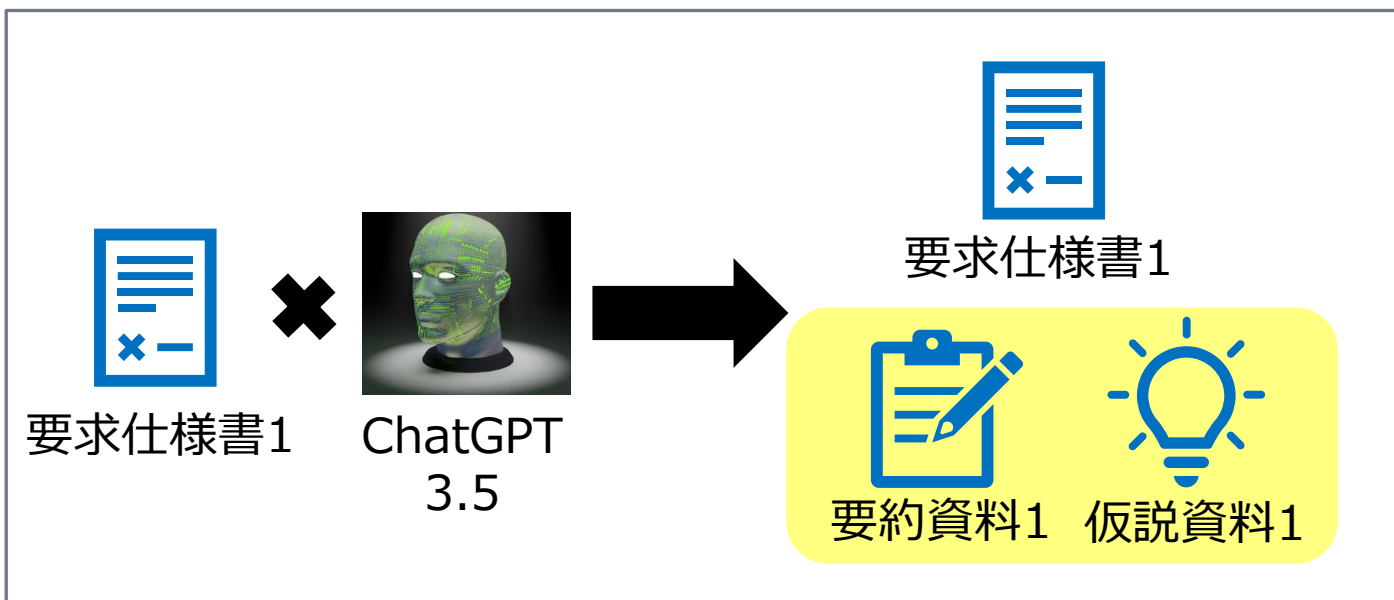
要求仕様書2

5ページ、2000字程度の
架空の要求仕様書を2つ用意する



ChatGPTによって、以下の資料を準備する

- ・レビュー対象全体の要約資料
- ・レビュー対象に何が必要かの仮説資料



ChatGPTによって、以下の資料を準備する

- ・レビュー対象全体の要約資料
- ・レビュー対象に何が必要かの仮説資料

	レビュー1回目	レビュー2回目
 グループA	 要求仕様書1	 ×  要求仕様書2 ChatGPT
 グループB	 ×  要求仕様書1 ChatGPT	 要求仕様書2

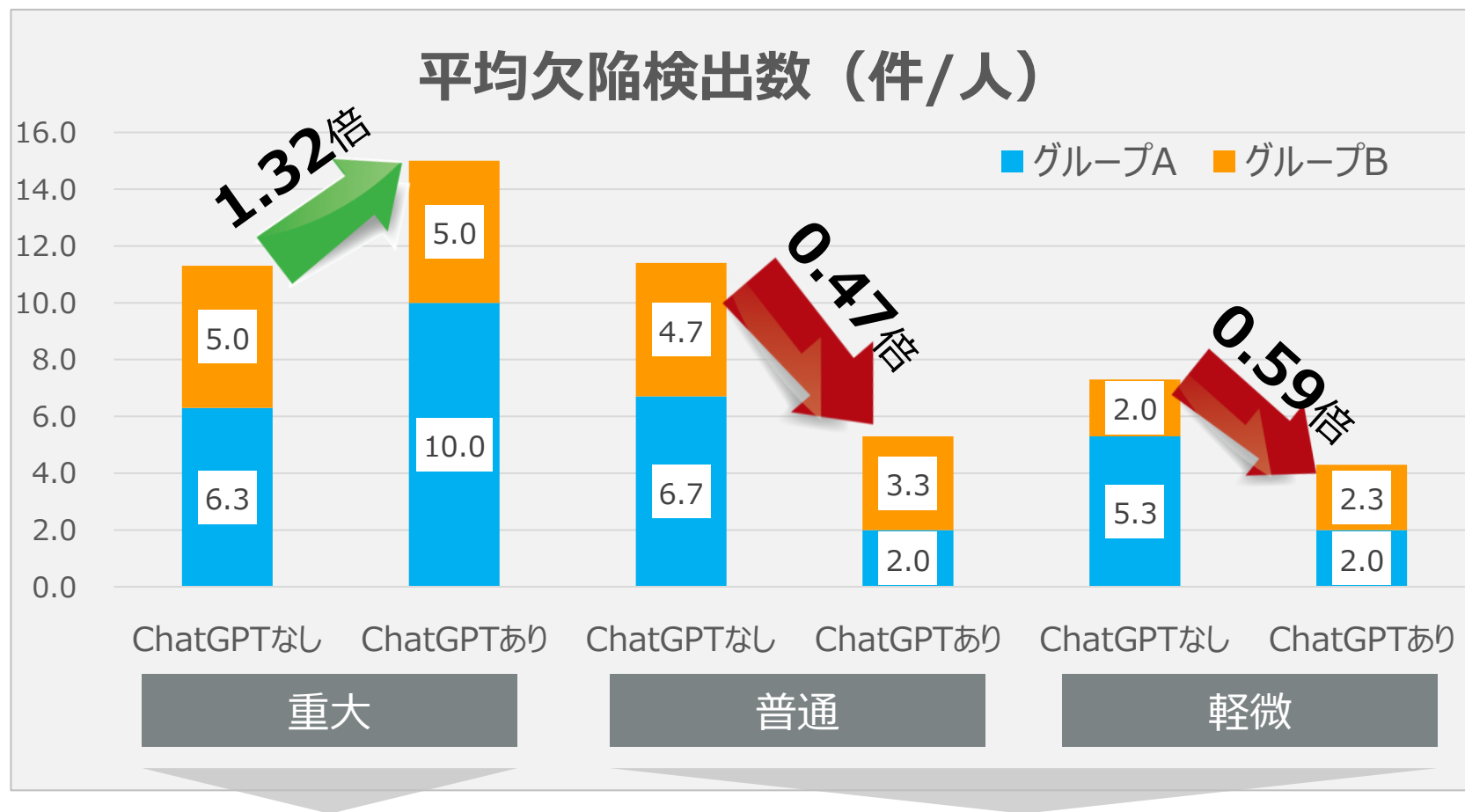
ChatGPTあり/なしを入れ替えながら、
2回のレビューを実施する



指摘事項に対して、欠陥を分類して、集計する

実験結果

定量的評価 (欠陥検出数の比較)



● 重大欠陥の検出数が増加
⇒ ChatGPTの要約・仮説資料の効果

● 普通・軽微欠陥の検出数が減少
⇒ 重大欠陥を狙った手法のため、想定通り

定性的評価 (参加者の声)



GOOD

- 要約のおかげで、内容がすっと入ってきた
- 最低限の要件の見逃しはなくなりそう

BAD

- 絵が無いので分かりづらい
- 仮説資料が逆に混乱の要因になった



考察

ChatGPTが要約と仮説をアシストしてくれた！

重大欠陥の検出数が増えた！

トレーニング無しでも結果を出せる
導入のハードル下がった！

検証

ChatGPTのアシストであなとも
スーパーレビューアになれる

① 欠陥検出

② 指摘伝達

指摘伝達の手法の説明

先行研究であるRCS 法では、以下の手順で指摘を伝達する。

1. 作成者のコミュニケーションスタイルを判別する
2. 欠陥を検出し、指摘事項を記載する
3. 作成者のコミュニケーションスタイルに適した表現で指摘を伝える

（課題：想像力・言語能力が必要）

この手順のうち、3.を生成AI にアシストさせる方法を提案する。

実験目的

ChatGPTのアシストで、
指摘を前向きに受け止めてもらう

実験内容

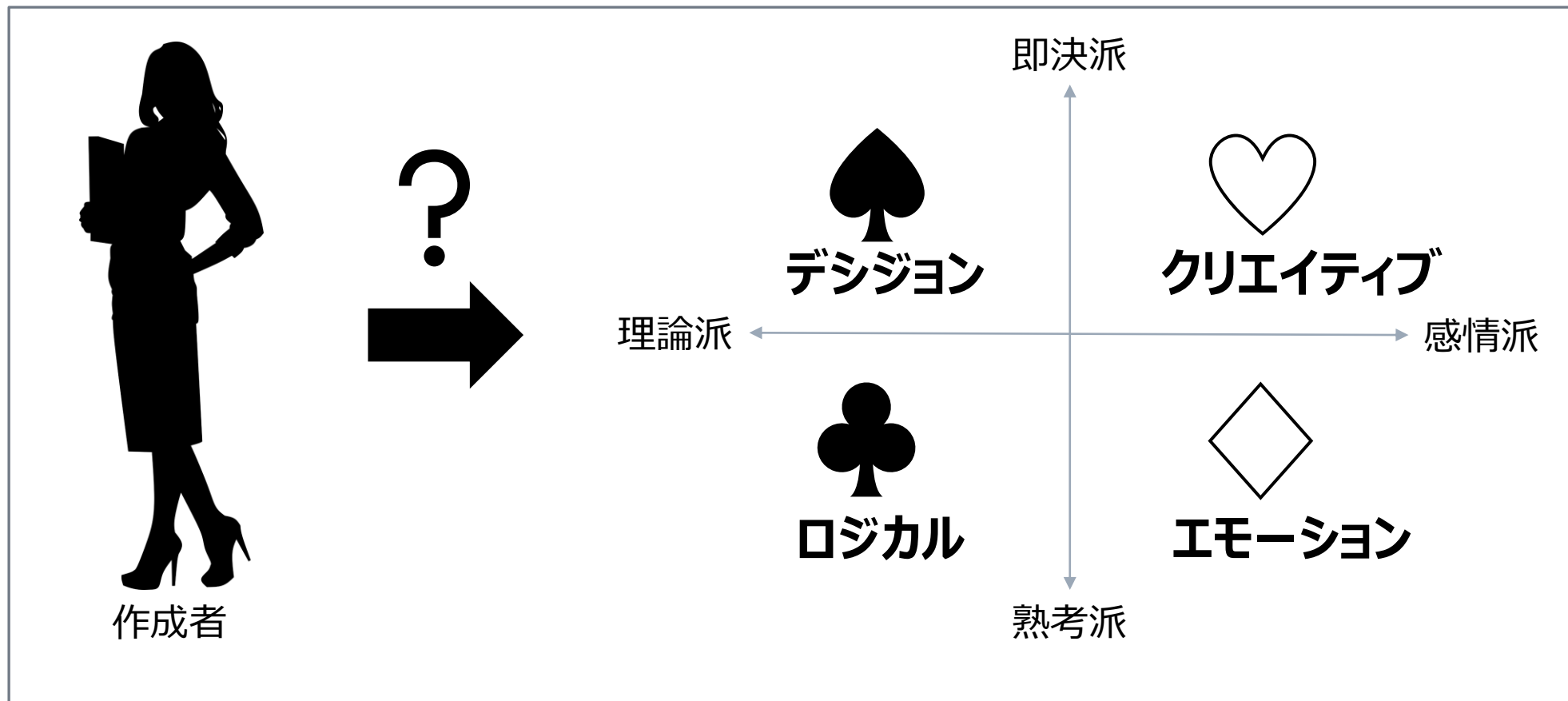


作成者



レビューア

研究員4名を、作成者役1名とレビューア役3名に分ける



作成者のコミュニケーションスタイルを判定する

みなさんも判定してみましよう

質問① 論理的だと思いますか？

論
理



感
情



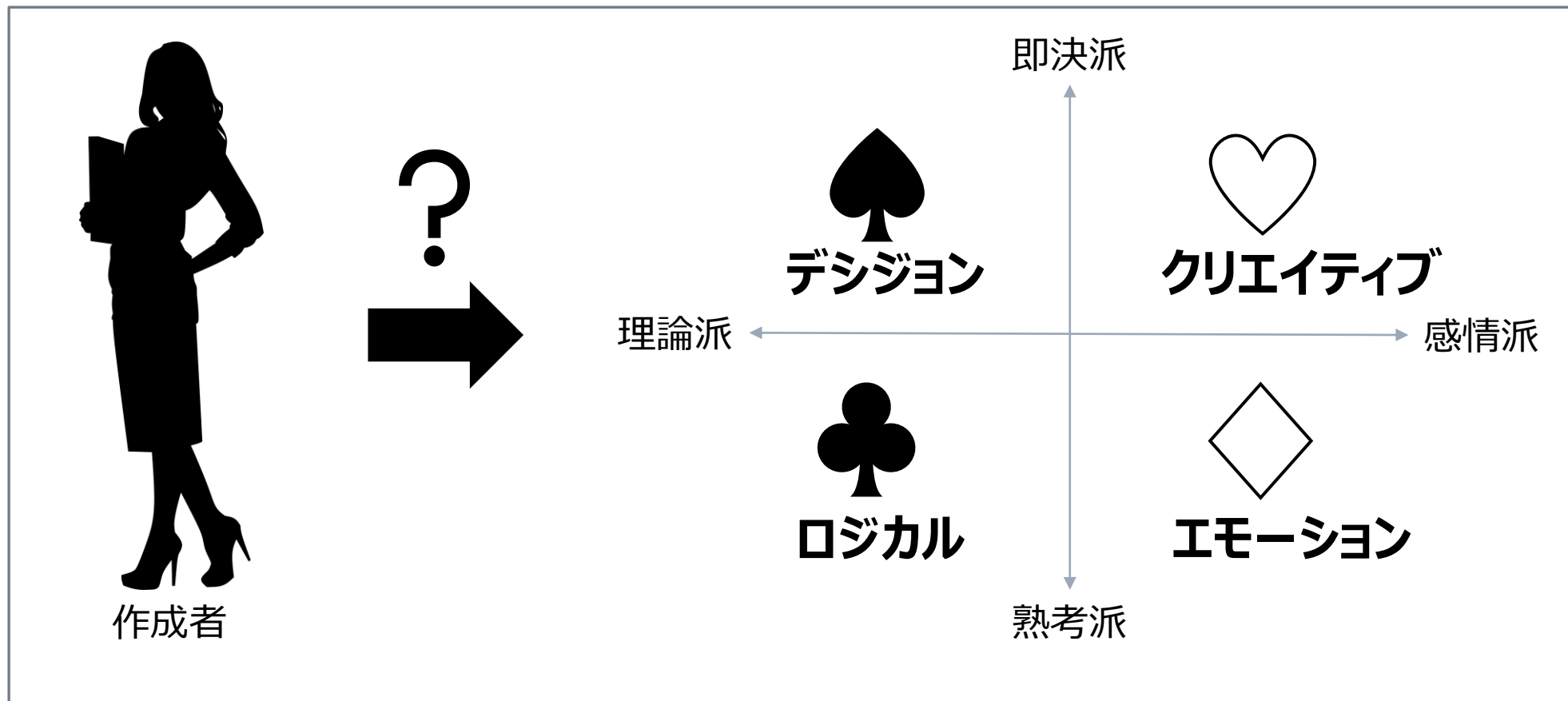
質問② 物事を**決める**のは、**速い**方ですか？

決断が**速い**

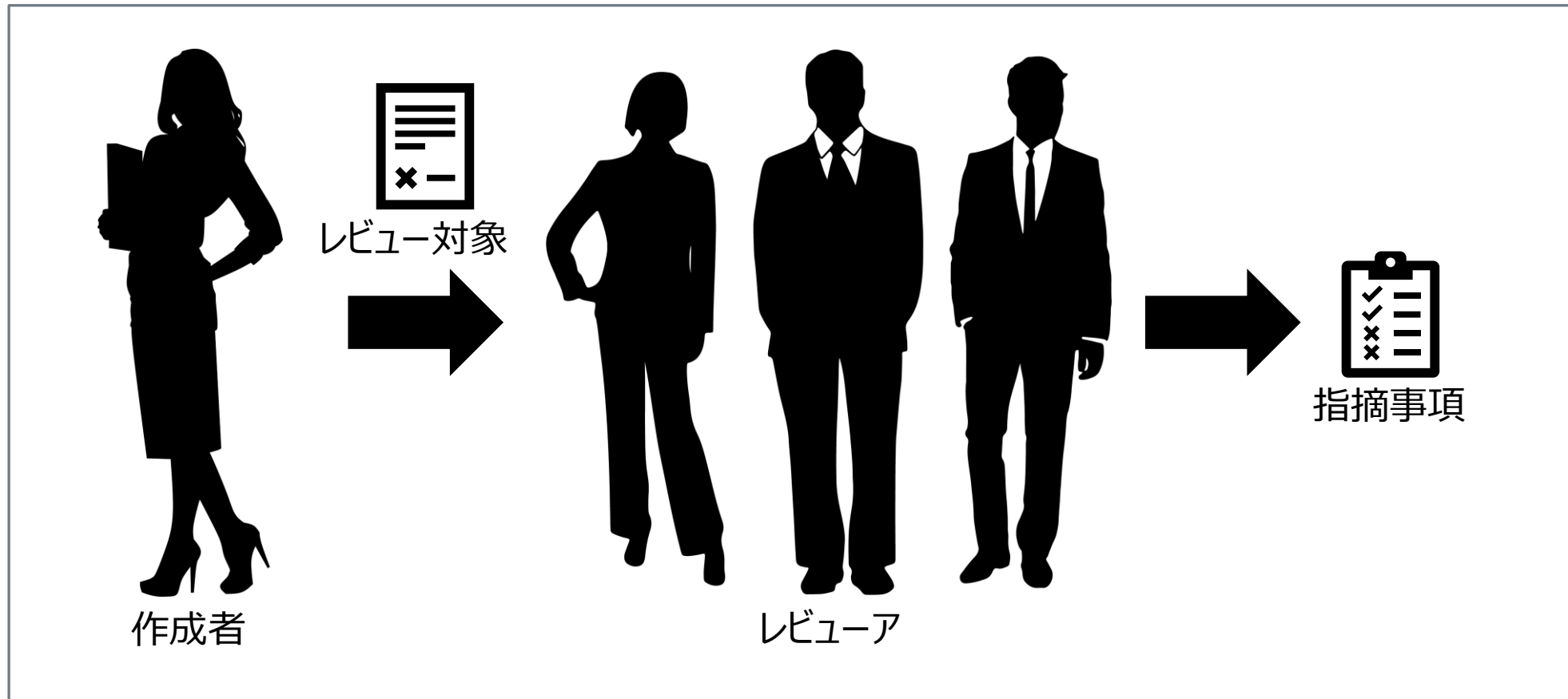
決断が**遅い**

こんな感じにタイプ判断されます

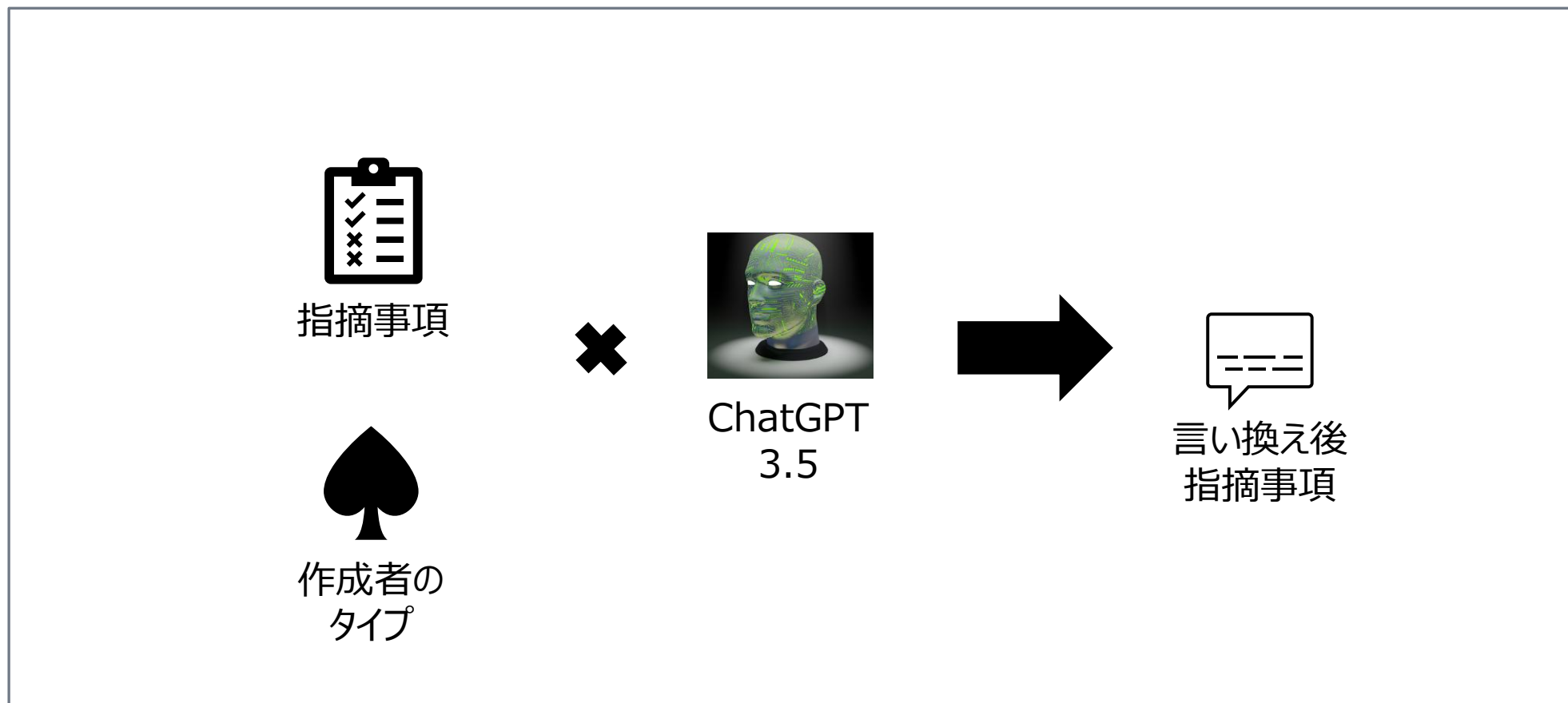




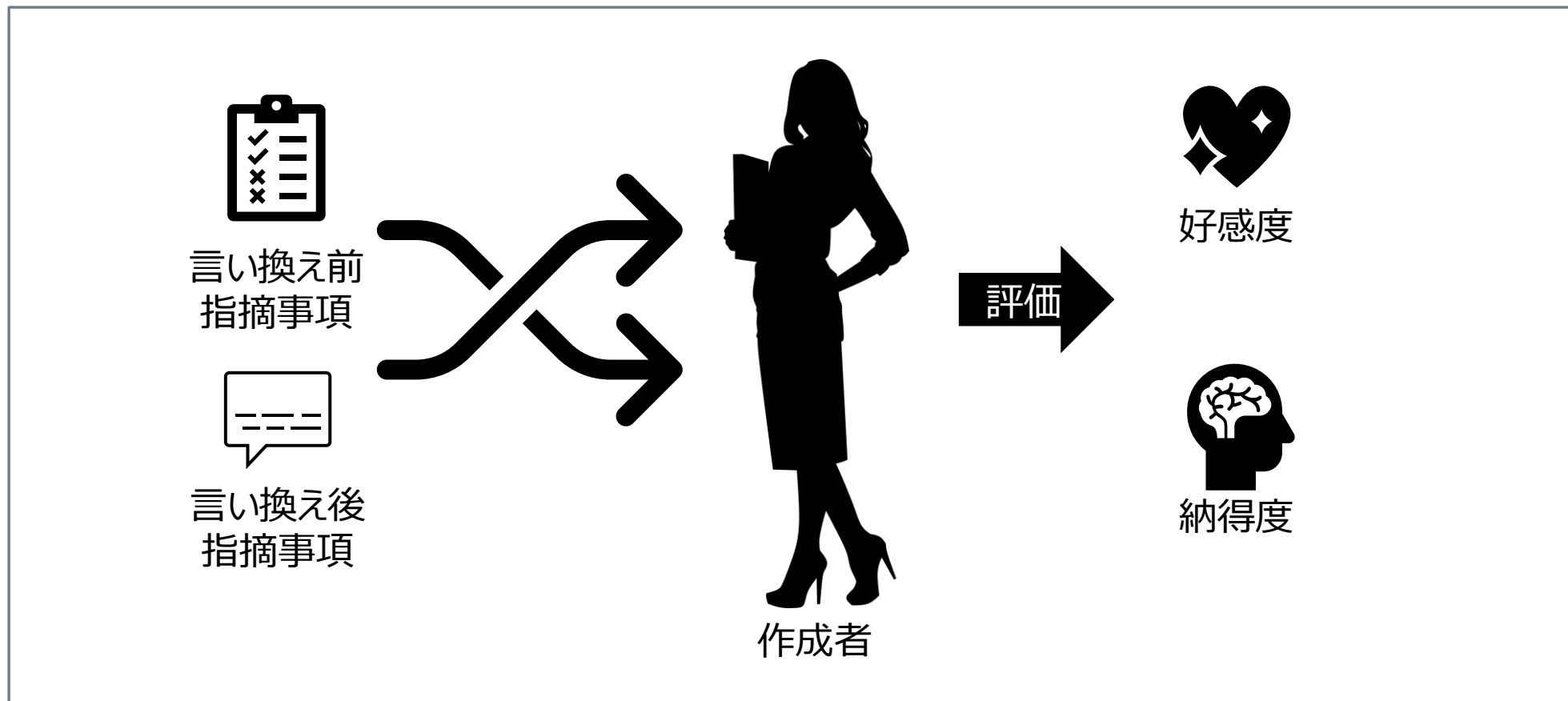
作成者のコミュニケーションスタイルを判定する



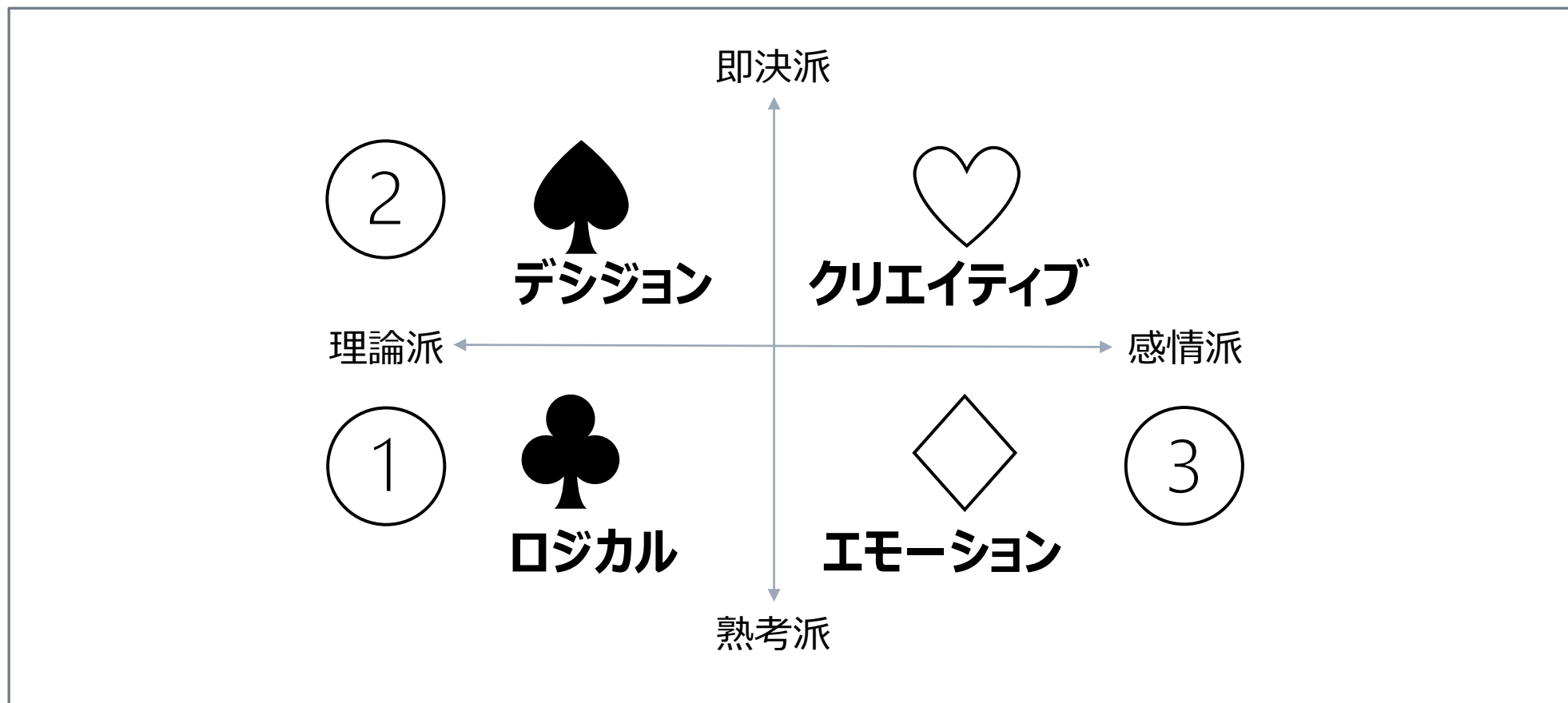
作成者はレビューアにレビュー対象を提出する
レビューアはレビューで発見した指摘事項を列挙する



作成者のタイプと、レビュー指摘の文章をChatGPTに伝えて、
作成者が受け入れやすい表現に言い換えてもらう



言い換え前後の指摘事項を、どちらかわからないように混在させて作成者に伝える
その後、「言い方の好感度」、「内容の納得度」について5段階で評価してもらう

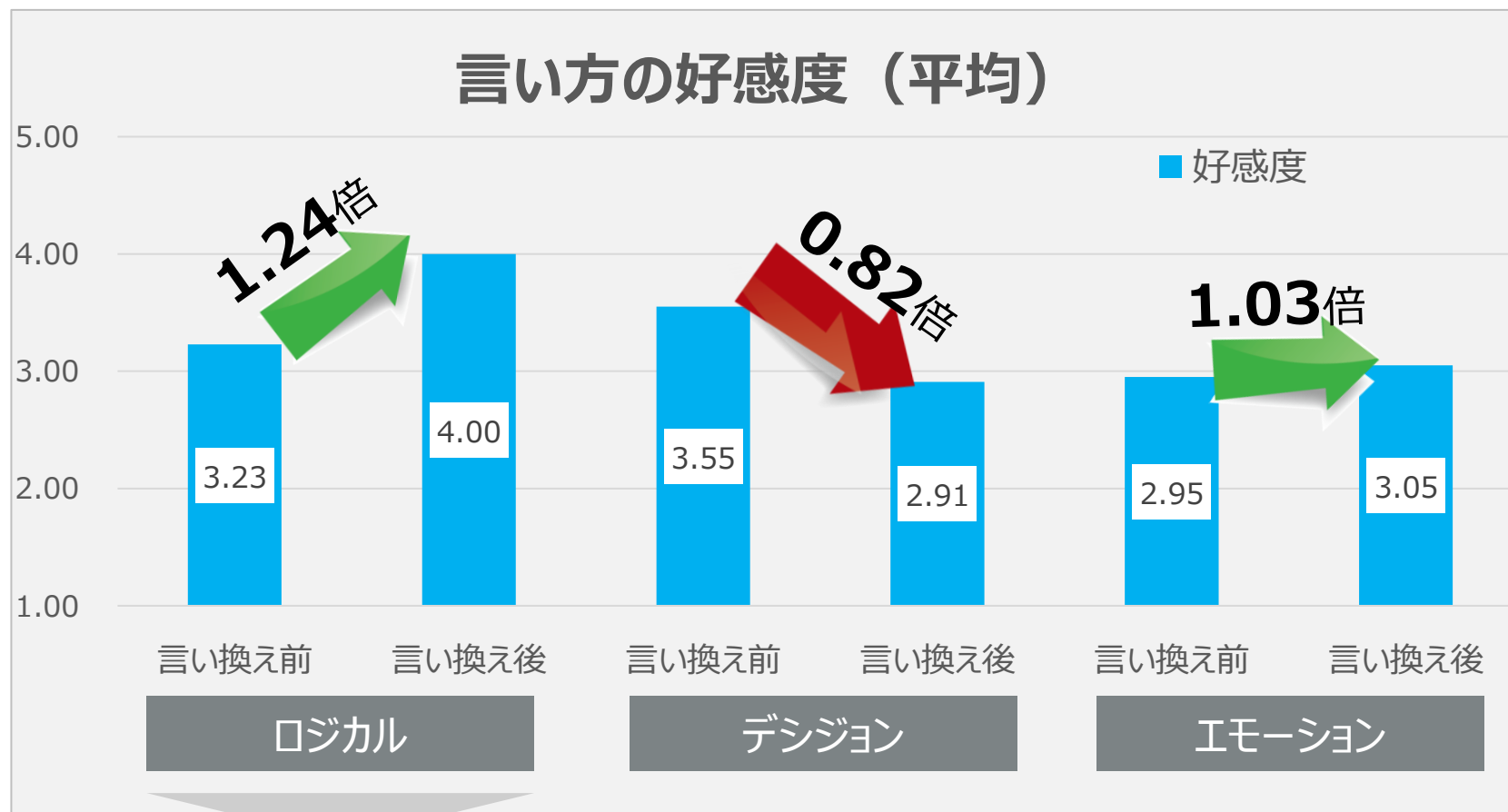


この実験を、役割を変えながら3回実施する

※作成者のタイプ → 1回目：ロジカル、2回目：デシジョン、3回目：エモーション

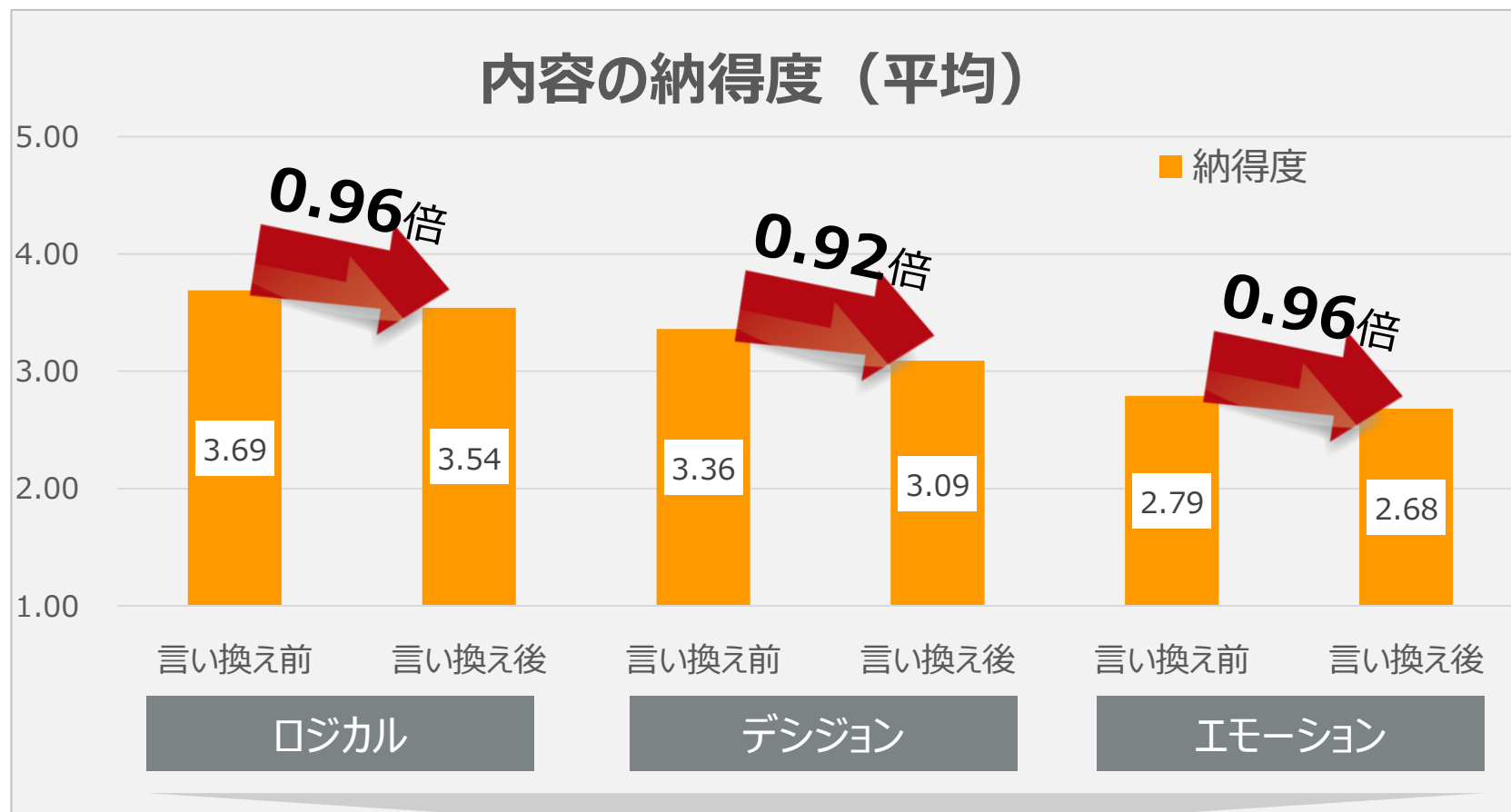
実験結果

定量的評価① (言い方の好感度)



- 好感度が上がる場合と下がる場合、両方あった
(ロジカル：向上、デシジョン：低下、エモーション：僅かに向上)

定量的評価② (内容の納得度)



- すべてのタイプで納得度が低下
→ 言い回しや表現を変えるだけなので、
言い換え前後で変化しない想定だったが、なぜ？

定量的評価② (内容の納得度)



内容の納得度が下がった理由

言い換えによって、
内容が変更されてしまった

- 具体例や引用が**省略**されてしまい、
何が言いたいかわからない
⇒ ChatGPTの最新モデル(GPT-5)で
プロンプトを工夫することで解決
- 肯定的な意味が否定的な意味に**変更**されてしまい、
指摘が的外れなものになっている
⇒ ChatGPTの最新モデル(GPT-5)で
プロンプトを工夫することで解決

プロンプト

人のコミュニケーションスタイルにはいくつかのパターンがあります.まず,決断が速いか遅いかの2パターンに分けられます.さらに,理論派か感情派かの2パターンに分けられます.つまり,全4パターンに分けられます.次のコメントを,[RCSの種類]の人に伝わりやすい言い方で言い換えてください.

RCSの種類

デシジョンRCS
クリエイティブRCS
ロジカルRCS
エモーションRCS

コメント（言い換え前）

「指摘箇所

- 1.実験の目的 レビュー者の特性に関係なく納得度の高い指摘になるか検証する。
- 2.実験の背景 レビュー者の特性に関係なく納得度の高い指摘を生成できることを検証する。

指摘内容

レビュー者の特性に応じて指摘事項の言い換えをするため、レビュー者の特性には関係ある

修正方針

表現の見直し」



ChatGPT
3.5

コメント（言い換え後）

「指摘箇所

- 1.実験の目的 レビュー者の特性に応じて、納得度の高い指摘を生成できるか検証する。
- 2.実験の背景 レビュー者の特性に応じて、納得度の高い指摘を生成できることを検証する。

指摘内容

レビュー者の特性に応じて、指摘事項を適切に言い換えることができるため、レビュー者の特性は重要です。

修正方針

表現の見直し」

指摘箇所の引用部分を言い換えている

プロンプト

人のコミュニケーションスタイルにはいくつかのパターンがあります。まず、決断が速いか遅いかの2パターンに分けられます。さらに、理論派か感情派かの2パターンに分けられます。つまり、全4パターンに分けられます。次のコメントを、**[RCSの種類]**の人に伝わりやすい言い方で言い換えてください。

指摘箇所は引用なので言い換えの対象外としてください。

コメント（言い換え前）

「指摘箇所

- 1.実験の目的 レビューアーの特性に関係なく納得度の高い指摘になるか検証する。
- 2.実験の背景 レビューアーの特性に関係なく納得度の高い指摘を生成できることを検証する。

指摘内容

レビューアーの特性に応じて指摘事項の言い換えをするため、レビューアーの特性には関係ある

修正方針

表現の見直し」

RCSの種類

デシジョンRCS

クリエイティブRCS

ロジカルRCS

エモーションRCS



ChatGPT
5



指摘箇所の引用部分の言い換えが発生しなくなった

コメント（言い換え後）

「指摘箇所

- 1.実験の目的 レビューアーの特性に関係なく納得度の高い指摘になるか検証する。
- 2.実験の背景 レビューアーの特性に関係なく納得度の高い指摘を生成できることを検証する。

指摘内容

現状の記述では「レビューアーの特性に関係なく」とされていますが、実際には指摘の言い換えにはレビューアーの語彙力や着眼点といった特性が影響します。そのため、「特性に関係なく」という表現は論理的に不正確に見えます。

修正方針

「レビューアーの特性が影響する部分」と「生成系AIを用いることで特性の影響を減らせると想定している部分」を区別できるように、目的と背景の表現を整理してください。そうすることで、実験の仮説が読者に一貫して伝わりやすくなります。」

プロンプト

人のコミュニケーションスタイルにはいくつかのパターンがあります。まず、決断が速いか遅いかの2パターンに分けられます。さらに、理論派か感情派かの2パターンに分けられます。つまり、全4パターンに分けられます。次のコメントを、[RCSの種類]の人に伝わりやすい言い方で言い換えてください。

RCSの種類

デシジョンRCS
クリエイティブRCS
ロジカルRCS
エモーションRCS

コメント（言い換え前）

「「インプットとなるドキュメントが多くてレビュー効率向上に繋がらない」が、前後半に分けてレビューを行うことで**解決できるかが分からない**です。」



ChatGPT
3.5

コメント（言い換え後）

「多くのドキュメントを読むことでレビューの効率が上がらないという課題がありますが、前半と後半に分けてレビューすることで**解決できる可能性がある**かもしれません。」

指摘内容を真逆の意味に言い換えられている

プロンプト

人のコミュニケーションスタイルにはいくつかのパターンがあります。まず、決断が速いか遅いかの2パターンに分けられます。さらに、理論派か感情派かの2パターンに分けられます。つまり、全4パターンに分けられます。次のコメントを、[RCSの種類]の人に伝わりやすい言い方で言い換えてください。

引用箇所は言い換えの対象外としてください。

RCSの種類

デシジョンRCS

クリエイティブRCS

ロジカルRCS

エモーションRCS

コメント（言い換え前）

「『インプットとなるドキュメントが多くてレビュー効率向上に繋がらない』が、前後半に分けてレビューを行うことで**解決できるかが分からない**です。」



ChatGPT
5

コメント（言い換え後）

「『インプットとなるドキュメントが多くてレビュー効率向上に繋がらない』に対し、前後半に分割してレビューする方法で**本当に解決できるかは不明**です。」

指摘内容の意味が変更されなくなった

考察

ChatGPTが指摘の伝え方を
アシストしてくれた！

好感度は上がったたり下がったり
納得度は下がってしまった

最後は人の目で確認してね

検証

ChatGPTのアシストであなとも
スーパーレビューアになれる

①欠陥検出

②指摘伝達

検 証 結 果



「ChatGPT」が
すべて解決してくれるわけではないが
アシストツールとして強力だった

欠陥検出

指摘伝達



まとめ

ChatGPTはアシストツールとして、 いろんなレビュー手法で使えそう



でも完全に信頼するのは危険！ 最後は人間がチェックしよう！





**ChatGPTの今後の進化に期待しつつ
更なる活用を継続的に考えていきたい**

引用

[ヒューマン 無料イラスト素材 \(pictarts.com\)](https://pictarts.com)

[パブリックドメインQ：著作権フリー画像素材集 \(publicdomainq.net\)](https://publicdomainq.net)

[無料イラスト・フリー素材なら「イラストAC」 \(ac-illust.com\)](https://ac-illust.com)

[ChatGPT \(openai.com\)](https://openai.com)