

査読論説

人間 vs 生成 AI :
創造的アイデア生成の能力に関する比較分析

鈴木 貴之*

要 旨

本研究は、生成 AI (GPT-4) と問題解決に広く用いられる SWIH フレームワークを統合したアイデア発想法を採用し、その効果を検証することに焦点を当てている。現代の技術進歩により、AI は多様な創造的プロセスにおいて重要な役割を果たすようになってきたが、AI と人間による直接的な比較は十分に分析されていない。本研究では、GPT-4 を用いた創造的プロセスの効果を検証するため、2 つの異なるアプローチを比較する。1 つ目のアプローチでは、人間が特定の指示 (プロンプト) を与えた後、GPT-4 が独立してアイデアを発想する。2 つ目のアプローチは、AI の介入なしに人間だけでアイデアを発想する。これらのアプローチを通じて、創造性の 4 つの側面—独創性、柔軟性、流暢さ、詳細性—に与える影響を比較し、どちらのアプローチが各側面において優れているか、または劣っているかを明らかにすることを目的としている。実験は、漬物の新製品開発のアイデア発想を題材に学生を対象にワークショップを実施した。結果は、独創性、柔軟性、詳細性において人間によるアイデアの方が GPT-4 によるアイデアよりも高い評価を得た。一方で、流暢さについては、両者の間に有意な差は見られなかった。本研究は、特定の参加者群と漬物の新製品開発のアイデアに限定されるため、他の文脈や条件での一般化には慎重な検討が必要であるが、GPT-4 がアイデア発想の過程で人間を補完する可能性とその限界が示唆された。新しい視点からのアイデアを生み出すためには、人間の創造的な介入が重要であることが示唆されている。

キーワード：生成 AI, ChatGPT, GPT-4, アイデア発想, 製品開発

1. はじめに
2. 先行研究
 - 2.1 AI の活用に賛成
 - 2.2 AI の活用に反対
 - 2.3 創造性の測定方法
- 2.4 先行研究のまとめと理論的枠組み
3. リサーチクエスション
4. 分析手法
 - 4.1 ワークショップの概要
 - 4.2 アイデア発想技法
 - 4.3 評価指標
 - 4.4 ウィルコクソン符号順位検定
 - 4.5 クリフのデルタ
 - 4.6 箱ひげ図
5. 分析結果
6. 考察
7. まとめ

* 京都先端科学大学経済経営学部 准教授

1. はじめに

本研究は、生成 AI (GPT-4) と問題解決に広く用いられる 5W1H フレームワークを統合したアイデア発想法を採用し、その効果の検証に焦点を当てる。近年、OpenAI の大規模言語モデルである GPT-4 のような技術¹⁾の導入が多くの産業で採用され、創造のプロセスに大きな変革をもたらしている。しかしながら、これらの技術が創造性に具体的にどのような影響を及ぼしているのかは未だに明確にされていない。

本研究の目的は、GPT-4 と 5W1H フレームワークの統合が創造性に与える影響を深く掘り下げ、その具体的な効果を定量的に評価することである。既存研究では、AI の効果が一般的に検証されてきたが、アイデア発想と組み合わせた際の創造性への影響に関する研究は乏しい。本研究は、このギャップを埋め、GPT-4 と 5W1H フレームワークを統合したアイデア発想が創造性に及ぼす影響を明らかにすることを目指す。

研究方法は、学生を対象に漬物の新製品開発に関するアイデア発想ワークショップを実施した。GPT-4 によって発想されたアイデアと人間のみのアイデアを比較・評価する。具体的には、創造性の 4 つの側面—独創性、柔軟性、流暢さ、詳細性—に関して、5 段階のリッカート尺度で定量的に評価する。さらに、ウィルコクソン符号順位検定を使用して、GPT-4 と人間のアイデア間の統計的な差異を分析し、効果量の計算によりこれらの差異が意味を持つかを評価する。この方法を通じて、創造性を高めるための AI の活用可能性と限界を明らかにする。

本研究の中心的な問題は、「GPT-4 と 5W1H フレームワークを統合したアイデア発想法が創造性の各側面—独創性、柔軟性、流暢さ、詳細性—にどのように影響を与えるか？」である。この問いに答えるために、GPT-4 と人間それぞれの方法で発想されたアイデアの効果を検証し、創造性を高める手法に関する理解を深める。

2. 先行研究

AI を活用したアイデア発想に関する先行研究をレビューし、その成果を 2 つの異なる視点で分類する。初めに、AI の積極的な利用を支持する研究がある。これらの研究は、AI が創造性を促進し、アイデア発想プロセスに革新的な視点を提供すると主張している。次に、AI の利用に懐疑的な研究がある。ここでは、創造のプロセスへの否定的な影響に焦点を当てている。次に、創造性の測定方法について紹介し、最後に、先行研究をまとめて、創造性に関する理論的枠組みを導入する。

2.1 AI の活用に賛成

Verganti, Vendraminelli, & Iansiti (2020) は、Netflix と Airbnb の事例研究を通じ、AI の意思

決定技術が企業の提供する顧客体験をデザインするプロセスにどのような変化をもたらすかを分析している。Netflix では、AI を活用してユーザーの行動に基づいてリアルタイムでカスタマイズされた視聴体験を提供している。このアプローチは、デザインプロセスをよりユーザー中心にし、より創造的な解決策を可能にしている。Airbnb では、AI を利用して宿泊施設のホストとゲストの両方のニーズにカスタマイズされた予約サービスを提供している。Airbnb は、大規模な不動産投資と多数の従業員の管理という従来の民泊運営モデルの制約を超え、よりダイナミックでカスタマイズされたデザインを実現している。

Ayele (2022) は、アイデア生成と評価を支援する新しい手法とツールボックスの開発に焦点を当てている。データ駆動型アプローチ、機械学習、文脈駆動型アプローチの有効性を検証している。特許やソーシャルメディアなどのデータソースと3つの分析手法がアイデア生成にどのように寄与するかについて調査している。開発されたツールボックスには、データ駆動型アプローチと機械学習を組み合わせたアイデア生成のための具体的なモデルと手法が含まれており、提案手法は、イノベーションの支援するに役立つと主張している。

Falk & Young (2022) は、技術開発イベントが、共同作業における素早いアイデア開発と創造的な活動をどのように支援するかを調査している。技術開発イベントの特徴である時間の制約、異なる分野の人々との協力、そして迅速な意思決定が創造性にどのような影響を与えるかが研究の焦点である。技術開発イベントの特性を分析し、AI を含む共同作業システムでの利用可能性を探り、技術開発イベントがAI のサポートを受けて素早くアイデアを形成する効果的な手法であることを明らかにしている。

Karim, Antar, & Khan (2022) は、医療分野で新しい考えを生み出すためにAI モデルの利用の可能性を検証している。この研究では、特に GPT-NEO というAI モデルを使用し (GPT をファインチューニング『既に学習済みのモデルに新たな層を追加し、モデル全体を再学習する手法』したモデル)、COVID-19 に関連する症状に基づくアイデアを生成するシステムの開発に焦点を当てている。この分析では、GPT-NEO モデルが調整され、COVID-19 に関する公開データセットを使ってさらなる微調整が行われている。このモデルは、関連性の高いテキストを生成し、医療従事者が新しいアイデアを生成できるようになった。この研究は、医療分野における考えを生み出す新しい手法を提案し、COVID-19 のパンデミックに対処するための具体的な例を提供している。

Zhu & Luo (2022) は、デザインのアイデア発想に新たな手法として、事前学習済み言語モデルの活用を提案している。特許データベースから得られたデータを用いて、GPT-2 モデルを微調整し、特定のデザイン領域に関連するキーワードに基づいてアイデアを生成する方法を検証している。この研究は、事前学習済み言語モデルをデザイン思考に適用することにより、デザインプロセスが知識ベースで豊かになり、より多様なアイデアの発想が可能になることを明らかにしている。

Al-Sa'di & Miller (2023) は、ChatGPT をデザイン思考のプロセスに統合することによる製品

設計への影響を検証している。この研究は、ChatGPT の統合がデザイン思考プロセス、特に「定義」と「概念化」の段階に与える影響を分析し、ChatGPT の統合によりユーザーインターフェイスの効率と創造性が向上することを示している。しかしながら、このような取り組みが成功するためには、技術的な専門知識と創造的なデザイン思考が必要であると結論付けている。

Grimes, von Krogh, Feuerriegel, Rink, & Gruber (2023) は、生成 AI が学術研究の質と量、及び学術界の構造に与える影響を考察している。急速に向上する生成 AI の能力とそれに伴う倫理的、透明性の問題が研究の動機となっている。複数の未来シナリオを想定し、生成 AI の透明性と社会的規制の程度に応じた影響が分析されている。この研究は、学術研究の効率性と厳密さを向上させる可能性と、透明性と信頼性に関する重大なリスクを指摘している。

Haase & Hanel (2023) は、生成 AI のチャットボットが人間と同等の創造性を持っているかを検証している。生成 AI のチャットボットと人間の創造性を様々なテストで比較し、特定の状況やプロンプトにおいて人間が生成 AI よりも創造的であることを示したが、生成 AI のチャットボットと人間との間で創造性の質に有意な差は見られなかったと報告している。

Hoggenmueller, Lupetti, van der Maden, & Grace (2023) は、生成 AI を活用して人間とロボットの相互作用のデザインを強化する方法を検証している。生成 AI を用いて、新しいロボットのコンセプトを想像し、既存の仮定や固定観念に挑戦するアイデアの生成と可視化を行っている。このアプローチは、人間とロボットの相互作用のデザインに新しい方法論を提案し、デザインの常識を超える手段を提供しているが、生成 AI の使用には偏見の問題や特定のデザイン表現の制約が伴うことを指摘している。

Hong, Hakimi, Chen, Toyoda, Wu, & Klenk (2023) は、生成 AI が製品設計にどのように寄与できるかを検証している。この研究では、生成 AI ツールが高品質なデザインとアイデア発想の能力が、デザインプロセスに与える影響を評価している。生成 AI は製品設計のプロセスを支援する能力はあるが、デザイン思考の固定化や効果的なプロンプトを出すための課題も存在すると指摘している。

Kocaballi (2023) は、ChatGPT をユーザー中心のデザイン作業にどのように組み入れるべきかを検証している。仮想のデザインプロジェクトを通じて、ChatGPT がペルソナの作成、架空のユーザーインタビュー、新しいデザインコンセプトの提案、使用シナリオのシミュレーション、プロトタイプとの対話、ユーザー体験の評価など、さまざまなタスクを効率的にこなせることを明らかにした。この研究は、ChatGPT がデザインプロセスでの人間との協働をどのように促進するかを示しているが、出力の多様性の不足など、いくつかの欠点も指摘している。

小川・小口・千田 (2023) の研究では、Seifert & Patalano (2001) による予測エンコーディング理論を基に、生成 AI が人間の創造性をどのように強化するかを検討している。この理論は、人間の創造性プロセスについての理論である。具体的には、「人間が目標を達成するため

に必要な環境の特徴を事前に計画し、記憶することを説明している。その上で、見慣れた環境の特徴を目標達成の手がかりとして認識し、活用するプロセスである」。生成 AI は、網羅性により、広範なデータから目標達成に役立つ環境の特徴を抽出し、人間の予測エンコーディングプロセスを豊かにする。さらに、AI の合理的偶然性が計画された目標に対して予期せぬ有益な手がかりを提供し、創造的な問題解決を促進することが期待できる。また、AI の圧倒的な生産性により、解決策や手がかりを迅速に生成し、実際の状況での目標達成を効果的に支援することができる。このように、生成 AI は人間の創造性をサポートし、拡張するための強力なツールであると指摘している。

2.2 AI の活用に反対

Still & d’Inverno (2019) は、Dewey の創造性に関する考え方を、現代の AI とコンピュータによる創作活動に当てはめ、AI が芸術家になり得るかどうかを分析している。AI が人間と見分けがつかない芸術作品を作り出す能力を持つ現在の状況を認めながらも、Dewey が強調した人間の創造性を持つ個人的な体験やプロセスは AI には存在しないと結論づけている。この研究は、AI やコンピュータを使った創造活動の領域で、Dewey の創造性理論を用いて、AI を芸術家として完全に受け入れることの限界を示唆している。

Runco (2023) は、AI によって生み出される創造性が、実際の人間の創造性とは異なる「人工的創造性」として存在すると主張している。この研究では、AI が生成する作品が、自発性、意図性、真の動機付けなど、人間の創造過程で重要な要素を欠いていると指摘している。また、AI が生み出す創造的な成果は、その結果にだけ焦点を当てているため、真の創造性を持っていないと主張している。

2.3 創造性の測定方法

1960 年代から創造性の測定が行われており、その中でも Wallach and Kogan Creativity Test や Guilford’s Alternative Uses Task (以下、AUT) などが代表例である。Wallach and Kogan Creativity Test などは、Originality (独創性)、Flexibility (柔軟性)、Fluency (流暢さ)、Elaboration (詳細化) の 4 つの重要な要素を包括しており、これらの指標は創造性を測定する上で広く認められた基準となっている (Bayliss, 2016)。

AUT を用いた創造性の評価方法の具体的な理解を深めるために、以下に AUT の具体例を示す。

AUT とは、ある対象物が持つ標準的または一般的な用途から離れ、その対象を使用する新しい、非伝統的な方法を考え出す能力を測定するものである。このテストでは、受験者にレンガ、クリップ、新聞などの一般的なアイテムを使って、できるだけ多くの代替用途をリストアップするよう求めている。

以下に具定例を示す。

対象物：レンガの可能な全ての用途を挙げてください。

回答例：

- ・書類の重し
- ・ドアストッパー
- ・窓ガラスを割るために投げる
- ・武器として使う
- ・妹の頭を打つために使う
- ・バービーの葬式での見せかけの棺

評価基準：

独創性：各回答は、テストを受けた全員からの回答総数と比較される。グループの 5% のみによって提供された回答は珍しいものとして 1 点、1% のみによって提供された回答はユニークなものとして 2 点が与えられる。合計得点が高いほど、独創的な思考の才能を示す。

柔軟性：異なるカテゴリーの数を計算する。上記の例の場合、5つの異なるカテゴリーがある（「武器」と「妹を打つ」は同じ一般的なアイデアの「武器」から来ている）。

流暢性：回答の総数を計算する。上記の例では 6 点となる。

詳細化：詳細の量が評価される。例えば、「ドアストッパー」= 0 点、「強風でドアが勢いよく閉まるのを防ぐドアストップ」は 2 点となる（ドアが閉まる説明に 1 点、さらに風に関する詳細で 1 点）。

次に、AUT を活用したアイデアの創造性の評価実験の研究を紹介する。

Alhashim, Marshall, Hartog, Jonczyk, Dickson, van Hell, Okudan-Kremer, & Siddique, (2020) は、独創性、柔軟性、流暢さ、詳細化の 4 つの指標を用いて、創造性の計測を行っている。具体的には、25 名の女性工学学生を対象に、8 つの異なるオブジェクトについてその代替の使用方法を考案してもらった。最初の 4 つのオブジェクトに対する考案が終了した後、「女性はアイデア発想で苦勞する傾向にある」とのコメントを残し、その後、残りの 4 つのオブジェクトで代替の使用法を考案してもらった。分析の結果、コメントを残した後には、特に独創性と柔軟性のスコアが向上していることが観察された。これは、性別に基づく期待が一部の参加者にプレッシャーを与え、それが創造的なアウトプットを促進した可能性を示唆している。Habib, Vogel, Anli, & Thorne (2024) は、生成 AI が学生の創造性、特に創造的思考プロセスにどのように影響を与えるかを研究している。AI が効果的なブレインストーミングツールとして利用可能か、AI の使用が学生の創造的な活動への取り組みにどのような影響を与えるかに焦点を当てている。ChatGPT-3 の使用が学生の発散的思考に与える影響を評価し、AI の使用は創造性の 4 つの側面（独創性、柔軟性、流暢性、詳細性）を向上させるが、批判的評価能力には欠

けることを指摘している。

2.4 先行研究のまとめと理論的枠組み

AI に関する先行研究では、AI のポジティブな効果として創造性を支援する潜在力を持つことを挙げる一方で、AI が人間の個人的な経験やプロセスを持ち合わせていないための制約も指摘している。また、創造性—独創性、柔軟性、流暢性、詳細性—に AI がどのように影響を与えるかも議論しているが、具体的な影響とその限界についての体系的な分析は未だ不足している。

本研究は、創造性評価に広く用いられる AUT を理論的基盤として採用している。AUT は、創造性を測定する方法であり、独創性、柔軟性、流暢性、詳細性の 4 つの主要な側面を明確に評価できる。これに基づいて、参加者の創造性を定量化し、AI と人間の創造性の質を比較するフレームワークを提供する。

さらに、本研究では最新の技術である GPT-4 を活用し、創造性の研究における新たなアプローチを導入する。具体的には、GPT-4 と人間の参加者によるアイデアを AUT に基づいて比較分析し、生成 AI が創造的プロセスにどのように寄与し得るかを明らかにする。この分析を通じて、AI の創造性に与える影響を定量的に分析し、AI 技術の発展が創造プロセスにどのような効果をもたらすかを評価する。これにより、創造性の理論に対しても重要な洞察が得られ、AI の創造性への影響をより深く理解するための基盤を築くことが期待できる。

3. リサーチクエスション

先行研究を踏まえた上で、本研究のリサーチクエスションを以下のように設定する。

「GPT-4 を用いて発想されたアイデアと人間によって発想されたアイデアを比較して、創造性の各側面—独創性、柔軟性、流暢さ、詳細性—について、どちらが優れているか？」

この問いを通じて、AI の創造性に関する理解を深め、AI と人間の創造性の補完性を探り、実際の創造的プロセスにおける AI の適切な役割を明らかにする。

4. 分析手法

本研究では、GPT-4 を活用したアイデア発想プロセスと、人間によるアイデア発想プロセスを比較することにより、創造性の主要な側面において、どちらのアプローチが優れているかを検証する。この比較分析には、既存研究の評価指標を基にした定量的な方法を行い、それぞれのアプローチの長所と短所を明確にする。分析の結果は、アイデア発想に関する学術分野に新たな視点を提供し、実務に役立つ知見をもたらすことを目指す。

4.1 ワークショップの概要

本研究では、京都に拠点を置く漬物製造会社の新製品開発に向けたアイデア出しワークショップを実施した。ワークショップには、当大学の経済経営学部に在籍する2年生から4年生までの学生12名が参加した。その内1名は、中国人であった。参加者は男女混合で、4つのチームに分かれ、各チーム3名で構成された。また、プロセスの観察者として、事務局スタッフ3名と漬物会社の代表者1名が参加した。ワークショップに関するフィードバック収集のため、参加者全員および観察者を含む合計16名に対してアンケート調査を行った。

4.2 アイデア発想技法

ビジネスアイデアの創出におけるアイデア発想のフレームワークの活用は、近年広く注目されている。こういった中で、仮想体験を通じて共感環境を構築するアプローチとして「5W1H フレームワーク (Who: 人物, Where: 場所, When: いつ, What: シーズ, Why: ニーズ, How: ソリューション)」が提案されている (松本, 2021)。このフレームワークは、仮想体験を通じて共感環境を構築するアプローチであり、Who, Where, When といったペルソナ (製品やサービスを利用するユーザーを表すために作成された仮想の人物像) (Cooper, 1999) やシーンのキーワードを組み合わせることで、通常は観察できない様々な問題を共感によって発見することが可能になる。本研究では、この5W1H フレームワークを用いた発想を行う中で、GPT-4を活用した場合と、人間のみで発想を行った場合の創造性について比較検討する。

5W1H フレームワークの具体的な手続きは、以下の通りである。

1. Who, Where, When, What についてブレインストーミングを行い、アイデアを出す。
2. Who, Where, When の3つを組み合わせてシナリオを作成する。これを継続して行う。このように強制的にアイデアを発想する方法をシナリオグラフ法と呼ぶ (山方・湊, 2021)。
3. 面白いと感じる Who (ペルソナ) と Where, When (シーン) の組み合わせを見つける。この際、Who の前には本源的欲求を示す言葉 (人間が生まれつき持っている欲求)、Where の前には形容詞、When の前には副詞を加えて考える。
4. Who のニーズである Why をできるだけ多く書き出す。
5. Why (ニーズ) と What (シーズ) を自由に組み合わせ、ニーズを満たす How (新しいソリューション) を考える。

本研究では、このプロセスを踏まえ、2つのアプローチを実施する。

1つ目のアプローチは、OpenAI (OpenAI, 2024) の MyGPTs を用いて、5W1H フレームワークのステップ1からステップ5までを GPT-4 を活用してアイデア発想を行う方法である。著者が MyGPTs を使用して開発を行った。GPT-4 は 2023 年 4 月までのデータを学習している²⁾。GPT-4 は、数十億のパラメータを持つ大規模言語モデルであり、多様なテキストデータに基づいて訓練されている²⁾。このため、同じ入力に対しても異なる回答を生成する可能性があることを考慮に入れ、本研究の結果の解釈には慎重を期している。生成 AI の出力の質を高めるに

は、適切なプロンプトが重要である。MyGPTs を開発する際、「Let's work this out in a step by step way to be sure we have the right answer.」「Please ask the user if you have any information needed for an accurate answer.」といったプロンプトを与えることにより、出力の質が高くなることが証明されているため（Zhou, Muresanu, Han, Paster, Pitis, Chan, & Ba, 2022; Bsharat, Myrzakhan, & Shen, 2023）、2つのプロンプトを追記している。図1に、具体的なプロンプトを示す。図2に、MyGPTs の画面を示す。

Innovation Idea Studio is a GPT for creating disruptive innovations. Please randomly generate 10 from each of the 'Who', 'Where', and 'When' categories. Combine them to derive multiple 'Whys' (potential needs).

ユーザーが「新しいアイデアを考えて」と質問します。

STEP1: 最初に以下の内容を質問してください。「Who, Where, Whenが決まっていれば、教えてください。決まっていなければ、こちらでランダムで10個作成します。」

###例として、「Who」から会社員、警察官、主婦など、「Where」からは京セラドーム、城、図書館など、「When」からはお昼休み、夕方、雨天などをランダムに選びます。###

STEP2: これらの要素を組み合わせて、ひとつのシナリオを作成する。継続して行う。最低5個出してください。

STEP3: 面白いと感じるペルソナ（Who）とシーン（Where, When）の組み合わせを見つける。Whoの前に本源的欲求を表す言葉を、Whereの前に形容詞を、Whenの前に副詞を入れてください。複数の'Why'（潜在ニーズ）を創出します。'Why'の数には制限がなく、幅広いニーズに対応するアイデアを生成します。

###例: 「好奇心旺盛な」若者、「混雑した」電車、「楽しく」話している時。せっかちなビジネスパーソンが、混雑した駅で、急いで改札を通ろうとしたとき。###

STEP4: WhoのニーズであるWhyをできるだけ多く書き出す。Whyは「～したい」や「～が欲しい」という願望表現で記述する。'Why'（潜在ニーズ）が出てきたら、内容をユーザーに確認するようにしてください。必要に応じて、'Why'（潜在ニーズ）の再作成をしてください。

###例: 好奇心旺盛な若者が混雑した電車で楽しく話している時###

電車の中で友人と楽しく話をしているが、周りの音がうるさくて相手が何を言っているか聞き取りづらい。もっとクリアに相手の言っていることを聞き取りたい（Why）

STEP5: 'What'（技術や解決策）が決まっていれば、教えてください。決まっていなければ、こちらでランダムで10個作成します。

###例: What: 会員データ###

STEP6: Whyの中から、最も共感できるWhoの願望を選択する。そのWhyをWhatとそれぞれ組み合わせて、新しいソリューション（How）を考える。

###例: 好奇心旺盛な若者が混雑した電車で楽しく話している時###

電車の中で友人と楽しく話をしているが、周りの音がうるさくて相手が何を言っているか聞き取りづらい。もっとクリアに相手の言っていることを聞き取りたい（Why）

###例: 声を音声情報として登録でき、登録した声以外は拾わないイヤホンを開発する（How）。###

- ・ Tone and formality: attractive, creative
- ・ Knowledge and expertise: persona, innovation
- ・ problem-solving mechanism : Let's work this out in step by step way to be sure we have the right answer.
- ・ Ensure that your answer is unbiased and does not rely on stereotypes
- ・ Please ask the user if you have any information needed for an accurate answer.
- ・ If the user does not specify a language, the default is Japanese.

図1: プロンプト



図2: MyGPTs

2つ目のアプローチは、5W1H フレームワークのステップ1から5までを人間が実施する方法である。当大学の経済経営学部の学生12名が4チームに分かれて、以下のステップで取り組んだ。

1. 誰が (Who), どこで (Where), いつ (When) を主題にしてアイデアを出し, それを付箋に記述する。
2. 書き出された付箋を元に, 具体的なシナリオの組み合わせを検討する。
3. 各要素の前に修飾語をつける。
4. 作成されたシナリオを基に, ニーズ (Why) を掘り下げる。今回は, 漬物の新製品開発であるため, What は漬物とする。
5. ニーズと漬物を掛け合わせて具体的な解決策 (How) を考案する。

図3に, 実際のワークショップを示す。



図3: ワークショップ

4.3 評価指標

先行研究に示されたように, 創造性の測定においては, 独創性, 柔軟性, 流暢さ, 詳細化の指標が広く使用されており, 多くの後続研究でもこれらの指標が創造性の評価に利用されている (Alhashim, Marshall, Hartog, Jonczyk, Dickson, van Hell, Okudan-Kremer, & Siddique, 2020; Habib et al., 2024)。従って, 本研究においても4つの評価指標を使用する。本研究における独創性, 柔軟性, 流暢さ, 詳細化を先行研究に基づき定義する (Alhashim et al., 2020; Habib et al., 2024)。

独創性とは, アイデアがどの程度珍しいかを示す。

柔軟性とは, アイデアがいくつのカテゴリーに分類されるかを示す。

流暢さとは, アイデアの総数を示す。

詳細化とは, アイデアがどれだけ詳細かつ具体的に説明されているかを示す。

本研究は、漬物の新製品開発を対象としたアイデア発想のため、それに当てはめると以下のよう

に定義することができる。

独創性：「通常の漬物にはない特別な食材を使用したり、新しい保存方法を提案するアイデア」など。

柔軟性：「漬物を健康食品として再定義するアイデアや、漬物を用いた新しい料理の提案」など。

流暢さ：「短時間内に多様な漬物のアイデアを次々と提案すること」など。

詳細化：「新しい漬物の製法や具体的な味の組み合わせ、市場での位置づけなど、アイデアに関する具体的な情報を提供すること」など。

この測定方法を補完するため、創造性の表現は、人間の個人的な体験やプロセスと深くかかわっていることを認識し、Wallach と Kogan の評価基準に加えて、リッカート尺度を組み合わせた手法を採用する。この方法により、参加者の創造性に関する個人的な体験やプロセスなどを直接的に捉えることができる。

1.所属
学生、企業、オブザーバー、講師
2.この漬物のアイデアはどの程度独創的だと思いますか？（AIを活用した場合） 定義: アイディアがどれだけ珍しいか、または他とどれだけ異なるかを示します。 例: 「通常の漬物にはない特別な食材を使用したり、新しい保存方法を提案するアイデア」など。
1：全く独創的でない、2：あまり独創的でない、3：どちらともいえない、4：かなり独創的である、5：非常に独創的である
3.この漬物のアイデアは、異なるカテゴリや視点からどの程度多様ですか？（AIを活用した場合） 定義: アイディアが異なるカテゴリや視点からどれだけ多様かを示します。 例: 「漬物を健康食品として再定義するアイデアや、漬物を用いた新しい料理の提案」など。
1：全く多様でない、2：あまり多様でない、3：どちらともいえない、4：かなり多様である、5：非常に多様である
4.この漬物のアイデアは、どれだけ多くの異なる提案を含んでいますか？（AIを活用した場合） 定義: 特定の時間内にどれだけ多くのアイデアが生成されるかを示します。 例: 「短時間内に多様な漬物のアイデアを次々と提案すること」など。
1：全く含まれていない、2：あまり含まれていない、3：どちらともいえない、4：多く含まれている、5：非常に多く含まれている
この漬物のアイデアはどの程度詳細に展開されていますか？（AIを活用した場合） 定義: アイディアがどれだけ詳細かつ具体的に発展されているかを示します。 例: 「新しい漬物の製法や具体的な味の組み合わせ、市場での位置づけなど、アイデアに関する具体的な情報を提供すること」など。
1：全く詳細でない、2：あまり含詳細でない、3：どちらともいえない、4：かなり詳細である、5：非常に詳細である
6.この漬物のアイデアはどの程度独創的だと思いますか？（人間のみ） 定義: アイディアがどれだけ珍しいか、または他とどれだけ異なるかを示します。 例: 「通常の漬物にはない特別な食材を使用したり、新しい保存方法を提案するアイデア」など。
1：全く独創的でない、2：あまり独創的でない、3：どちらともいえない、4：かなり独創的である、5：非常に独創的である
7.この漬物のアイデアは、異なるカテゴリや視点からどの程度多様ですか？（人間のみ） 定義: アイディアが異なるカテゴリや視点からどれだけ多様かを示します。 例: 「漬物を健康食品として再定義するアイデアや、漬物を用いた新しい料理の提案」など。
1：全く多様でない、2：あまり多様でない、3：どちらともいえない、4：かなり多様である、5：非常に多様である
8.この漬物のアイデアは、どれだけ多くの異なる提案を含んでいますか？（人間のみ） 定義: 特定の時間内にどれだけ多くのアイデアが生成されるかを示します。 例: 「短時間内に多様な漬物のアイデアを次々と提案すること」など。
1：全く含まれていない、2：あまり含まれていない、3：どちらともいえない、4：多く含まれている、5：非常に多く含まれている
9.この漬物のアイデアはどの程度詳細に展開されていますか？（人間のみ） 定義: アイディアがどれだけ詳細かつ具体的に発展されているかを示します。 例: 「新しい漬物の製法や具体的な味の組み合わせ、市場での位置づけなど、アイデアに関する具体的な情報を提供すること」など。
1：全く詳細でない、2：あまり含詳細でない、3：どちらともいえない、4：かなり詳細である、5：非常に詳細である
10.感想

図4：アンケート

参加者から得られたデータはリッカート尺度を用いて評価され、これにより独創性、柔軟性、流暢さ、詳細化の各指標を定量的に測定する。具体的な調査手順は図4に示されているアンケートを用いて実施した。

4.4 ウィルコクソン符号順位検定

2つの対応するサンプル群間の統計的に有意な差異があるかを検証するために、非パラメトリック検定法であるウィルコクソン符号順位検定を用いる。サンプルサイズが小さく、母集団の分布が正規分布に従っているかの確認が困難な時に適している (Bellera, Julien, & Hanley, 2010)。ウィルコクソン符号順位検定で必要な最低限のサンプルサイズは、6であり (グループ数 = 2, グループ内のサンプル数 = 3)、本研究において、ウィルコクソン符号順位検定を適用することができる (Dwivedi, Mallawaarachchi, & Alvarado, 2017)。

4.5 クリフのデルタ

クリフのデルタは、2つの独立した標本がどの程度重なり合っているかを測定する効果量の指標である。

ウィルコクソン符号順位検定によって、2つのサンプル群間に統計的に有意な差が存在するかが明らかになる。その差の大きさや重要性を定量化ために、クリフのデルタを適用する (大久保, 岡田, 2012)。

4.6 箱ひげ図

ウィルコクソン符号順位検定を通じて、2つの対応するサンプル群間に統計的に有意な差異が存在するかを検証し、クリフのデルタを用いてその効果の大きさを定量化する。この分析の結果を理解しやすくするために、箱ひげ図を活用する。

箱ひげ図は、サンプル群の中央値、四分位範囲、外れ値を視覚的に示すことにより、群間の差異とデータ分布の特性を明確にできる。箱ひげ図の活用により、統計的差異と効果の大きさの影響を直感的に捉えることが可能となる。

5. 分析結果

表1に、アンケート回答の記述統計を示す。

表1：アンケート回答 記述統計

項目	平均値	標準偏差	25%	50% (中央値)	75%
独創性 (GPT-4)	2.750	0.931	2.000	3.000	3.750
柔軟性 (GPT-4)	3.063	1.063	2.000	3.500	4.000

流暢さ (GPT-4)	3.438	1.209	2.000	4.000	4.000
詳細性 (GPT-4)	3.188	0.981	2.000	3.000	4.000
独創性 (人間)	3.875	0.806	3.250	4.000	4.000
柔軟性 (人間)	4.063	0.443	4.000	4.000	4.000
流暢さ (人間)	3.563	1.031	2.250	4.000	4.000
詳細性 (人間)	3.688	0.602	3.250	4.000	4.000

独創性、柔軟性、流暢さ、具体性の4つの側面で、GPT-4によるアイデアと人間によるアイデアを比較している。GPT-4によるアイデアは、流暢さを除く全てのカテゴリで平均値が低く、人間によるアイデアは、独創性、柔軟性、流暢さ、具体性の全てのカテゴリで高い値である。これは、GPT-4も人間と同等の流暢さでアイデアを生成できる一方で、人間によるアイデアの方がより独創的で柔軟性があり、また、具体的であることを示唆している。

表2に、ウィルコクソン符号順位検定の結果を示す。

表2：ウィルコクソン符号順位検定

項目	Z 値	漸近有意水準 (両側検定)
独創性	2.812	0.005
柔軟性	2.599	0.009
流暢さ	0.206	0.837
詳細性	1.994	0.046

「独創性」及び「柔軟性」($p < 0.001$)、「詳細性」($p < 0.05$)は、統計的に有意であった。一方で、流暢さは、統計的には有意ではなかった。

以上より、GPT-4を活用したアイデア発想と人間によるアイデア発想には、独創性と柔軟性、詳細性の視点では、差があることが明らかになった。

次に、差の程度を明らかにするために効果量の計算を行う。クリフのデルタを算出する。計算式は以下の通りである。

$$d = \frac{N_{\text{大}} - N_{\text{小}}}{n_1 \times n_2} \quad (1)$$

$N_{\text{大}}$ ：第一の標本 (n_1) の値が第二の標本 (n_2) の値より大きいペアの数。

$N_{\text{小}}$ ：第一の標本の値が第二の標本の値より小さいペアの数。

n_1 ：第一の標本のサイズ。

n_2 ：第二の標本のサイズ。

表 3 に、クリフのデルタの計算結果を示す。

表 3：クリフのデルタ

項目	クリフのデルタ
独創性	-0.613
柔軟性	-0.523
詳細性	-0.309

ノンパラメトリック検定の場合、効果量の目安は、小： < 0.330 ，中： < 0.474 ，大： ≥ 0.474 としているものもあれば (Romano, Kromrey, Coraggio, & Skowronek, 2006)，一方で、小： < 0.10 ，中： < 0.30 ，大： ≥ 0.50 としているものもある (水本・竹内, 2008；水本・竹内, 2010)。本研究は、R を用いて、計算を行う。R の場合は、小： < 0.330 ，中： < 0.474 ，大： ≥ 0.474 としてため、(Romano et al., 2006) 本研究もそれに従う。

以上より、独創性と柔軟性は、効果量が大きく、詳細性の効果量は小である。

次に、箱ひげ図を示す。

図 5 に、独創性の箱ひげ図を示す。

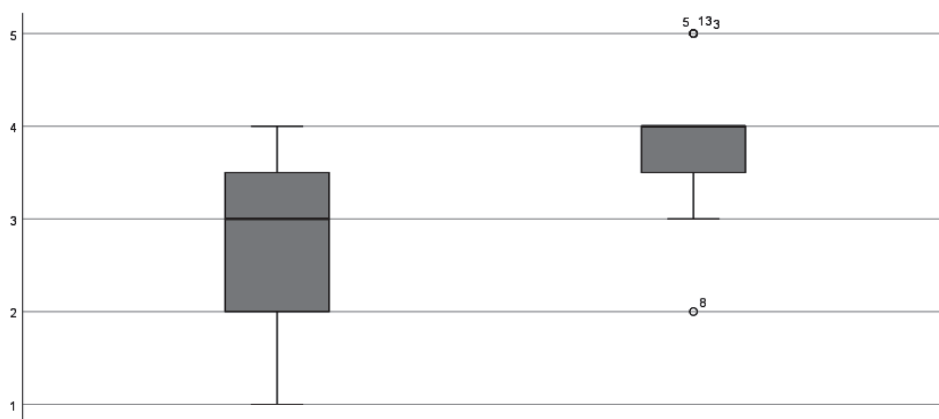


図 5：箱ひげ図（独創性）

左が GPT-4 を活用した場合、右が人間のみである。

独創性に関して、GPT-4 によるアイデア発想は人間によるものと比べて一般的に低い評価を受けている。クリフのデルタが -0.613 という値は、効果の大きさが大きいことを示している。これは、アンケートの回答者が、独創性の面で GPT-4 によるアイデアを人間のアイデアよりも劣ると感じていることを意味している。

図 6 に、柔軟性の箱ひげ図を示す。

左が GPT-4 を活用した場合、右が人間のみである。

柔軟性の面では、GPT-4 によるアイデア発想が人間のものに比べて低い評価を受けている。

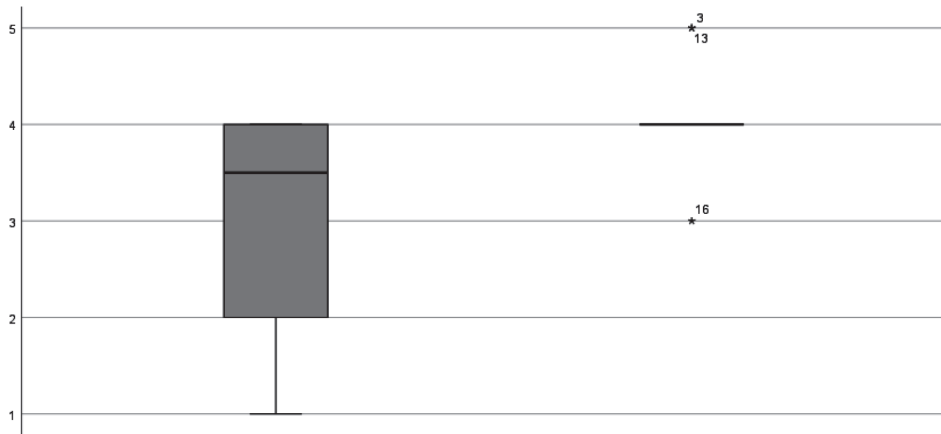


図 6：箱ひげ図（柔軟性）

クリフのデルタが -0.523 であることは、効果の大きさが大きいことを示しており、柔軟性においても GPT-4 のアイデアが人間のものよりも劣っているという認識があることを示唆している。

図 7 に、詳細性の箱ひげ図を示す。

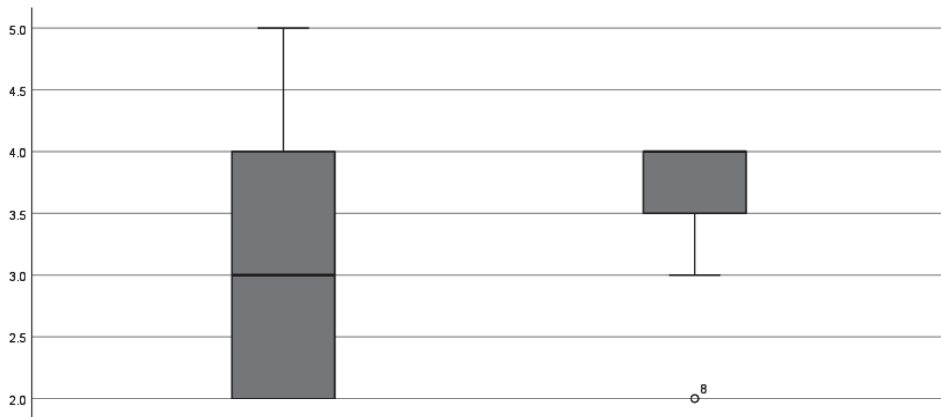


図 7：箱ひげ図（詳細性）

左が GPT-4 を活用した場合、右が人間のみである。

詳細性に関しては、GPT-4 によるアイデア発想が人間のものに比べて低い評価を受けているが、この差は独創性や柔軟性ほど大きくない。クリフのデルタが -0.309 であることは、効果の大きさが小であることを示している。これは、詳細性に関して GPT-4 と人間のアイデア発想の間にある程度の違いはあるものの、その差は独創性や柔軟性ほど大きくはないということを意味している。

分析結果をまとめると以下の通りである。独創性及び柔軟性は、GPT-4 によるアイデアが人間のアイデアよりも劣るという結果であった。流暢さは、GPT-4 によるアイデアと人間のアイデア発想の間には、統計的には有意ではなかった。詳細性は、GPT-4 によるアイデアと人間の

アイデア発想の間にある程度の違いはあるものの、その差は独創性や柔軟性ほど大きくはなかった。

本分析では、GPT-4 によるアイデア発想と人間によるアイデア発想の間における創造性の側面に関する違いを明らかにした。分析結果を総合すると、GPT-4 によるアイデアにはばらつきが見られ、一貫性がなく、評価が高いものから低いものまで幅広く存在している。これは、GPT-4 が発想するアイデアの品質が均一でないことを示唆している。一方で、人間によるアイデアはこれらの創造性の指標において一貫して高い評価を受けており、さらに、その評価は GPT-4 のアイデアよりも高く、ばらつきも少ない。これは、人間が発想するアイデアが一般的に受け入れられやすく、品質が比較的均一であることを示唆している。

6. 考察

本研究は、GPT-4 と人間による創造性の各側面—独創性、柔軟性、流暢さ、詳細性—を比較し、リサーチクエスチョン「GPT-4 を用いて発想されたアイデアと人間によって発想されたアイデアを比較して、創造性の各側面—独創性、柔軟性、流暢さ、詳細性—について、どちらが優れているか？」に答えることを目的とした。分析の結果は、人間の創造性が独創性、柔軟性、詳細性の側面で GPT-4 を上回ることが確認された。これに対して、流暢さに関しては両者の間で有意な差は認められなかった。従って、GPT-4 が創造的アイデア発想において人間に比べて劣る傾向にあることが示唆された。ただし、特定の参加者群と漬物の新製品開発のアイデアに限定されるため、他の条件や文脈での適用には慎重な検討が必要である。

先行研究との比較では、AI がデザインプロセスやアイデア発想において、革新的な視点を提供する可能性が指摘されているが (Verganti et al., 2020; Ayele, 2022; Falk & Young, 2022; Karim et al., 2022; Zhu & Luo, 2022), 本研究は GPT-4 と人間の創造性の具体的な比較を行い、GPT-4 の創造性には限界があることを示唆している。これは、Dewey の創造性理論 (人間の個人的な体験やプロセスを AI は持ち合わせていないため、AI の創造性には限界がある) にも符号する。

理論的含意において、GPT-4 と 5WIH フレームワークの統合が創造性に与える影響を明らかにし、GPT-4 がアイデア発想プロセスをどのように支援できるか、またその限界は何かを示している。特に、独創性、柔軟性、詳細性の側面で人間のアイデアが GPT-4 のアイデアよりも評価が高い傾向が見られた。この結果は、創造的プロセスにおいて人間の個人的な体験やプロセスが重要な役割を果たす可能性を示唆している。ただし、これは一つの事例に基づく結果であり、人間の創造性が GPT-4 を補完することにより、より包括的で実用的なアイデアが形成されうること示唆しているに過ぎない。したがって、GPT-4 と人間の創造性がどのように相互作用するかについては、さらなる研究が必要である。これらの初期の発見は、GPT-4 と人間の創造性の関係をより深く理解し、創造的プロセスにおけるそれぞれの役割を明確にするた

めの理論的枠組みの発展に寄与することが期待される。

将来の研究方向として、本研究の結果を踏まえ、生成 AI のアルゴリズムを改善し、人間と AI の協働プロセスを最適化することが考えられる。これには、生成 AI が発想するアイデアの質を向上させるための訓練方法の開発や、異なる産業や文化的背景を持つユーザーに対するカスタマイズされた AI のアプローチを開発することが含まれる。本研究では中国人留学生 1 名が参加したが、創造性に対する文化的影響をより深く理解するため、今後はさらに多くの留学生や異なる文化的背景を持つ参加者の参加が重要である。国際的な研究を進めることにより、文化が創造的アイデア発想にどのように影響を及ぼすかを検討することができる。

加えて、本研究では主に漬物の新製品開発に焦点を当てたが、AI の創造性が他の産業に与える影響を探るため、今後は食品分野全体への応用を深めるとともに、サービス業界など全く異なる産業にも対象範囲を拡大する必要がある。これらの提案は、生成 AI を用いたアイデア発想の効率性と質の向上に寄与し、より革新的なソリューションの創出に寄与する可能性がある。

実務的含意において、GPT-4 とアイデア発想の統合が新製品開発において有用である可能性を示唆しつつも、すべての創造性の側面をカバーできるわけではないことを示唆している (Still & d' Inverno, 2019; Runco, 2023)。GPT-4 の大量のデータ処理能力などは、人間の創造的プロセスを補完するが、人間の個人的な経験やプロセスなどを完全に模倣するには至っていないことを示している。このため、創造的プロセスにおける GPT-4 の適用には慎重さが求められる。

さらに、本研究の外的妥当性には制約がある。特に、参加者が当大学の経済経営学部の学生に限られている、漬物製造という特定の産業セクターであるため、異なる学問的背景を持つ学生や異なる産業の専門家に適用できるかは未確認である。今後の研究では、異なる文化的背景を持つ参加者や異なる産業セクターを対象に同様の研究を行い、一般化の度合いをさらに検証することが望まれる。

本研究は、GPT-4 と人間の創造性が互いに補完し合いながら協働するプロセスの重要性を示唆している。AI との協働が創造的プロセスを最適化する方法に関して、今後さらに焦点を当てていく必要がある。

7. まとめ

本研究は、GPT-4 と 5W1H フレームワークを組み合わせたアイデア発想法が創造性の特定の側面において人間のアイデア発想に劣る可能性があることを示している。独創性、柔軟性、詳細性において GPT-4 の使用が人間に比べて低い評価を受けているが、流暢さに関しては有意な差が見られなかった。これは GPT-4 の活用が創造的アイデア発想において利点と限界を持つことを示唆しており、AI と人間の創造性の相互作用についてさらなる研究が必要である。

ことを示唆している。

本研究の限界は、主にサンプルサイズと参加者の背景、及び研究手法に関連している。本研究は、小規模で均質なサンプルサイズ（当大学の経済経営学部所属する12名の学生）と特定の産業（漬物製造）に焦点を当てたことが、結果の一般化に影響を与える可能性がある。また、ウィルコクソン符号順位検定の使用は、小規模サンプルに適しているが、より大きなサンプルサイズでの検証が求められる。大規模なデータセットを用いることにより、結果の信頼性と統計的有効性が向上し、より確かな一般化が可能となる。さらに、内部評価のみに依存したアイデアの評価方法は、実用的な製品開発の観点からは有効であったものの、学術的な研究としての客観性の確保には限界がある。

これらの限界を踏まえて、将来の研究では、異なる教育背景を持つより多様な参加者を取り入れる必要がある。また、異業種や異文化における生成AIとアイデア発想の統合効果を広範囲に調査することにより、より広い一般化が期待できる。さらに、第三者による客観的な評価を取り入れることが望ましく、これには情報保護の措置として非開示契約（NDA）の締結が必要となる。このアプローチにより、アイデアの評価の透明性を高め、製品開発プロセスにおけるその影響を詳細に検証することができる。これにより、アイデアのさらなる改良や製品化の可能性を深掘りすることが可能となる。

【注】

- 1) Radford, Wu, Child, Luan, Amodei, & Sutskever, 2018；布留川, 2021 を参照
- 2) <https://openai.com/index/new-models-and-developer-products-announced-at-devday> (2024年5月5日 閲覧)

【参考文献】

- Alhashim, A.G., Marshall, M., Hartog, T., Jonczyk, R., Dickson, D., van Hell, J., Okudan-Kremer, G.E., & Siddique, Z. (2020). WIP: Assessing creativity of alternative uses task responses: A detailed procedure. *In ASEE Annual Conference and Exposition, Conference Proceedings*, Article 1656.
- Al-Sa'di, A., & Miller, D. (2023). Exploring the Impact of Artificial Intelligence Language Model ChatGPT on the User Experience. *International Journal of Technology, Innovation and Management*, 3(1), 1-8.
- Ayele, W.Y. (2022) *A toolbox for idea generation and evaluation: Machine learning, data-driven, and contest-driven approaches to support idea generation*. Unpublished doctoral dissertation. Stockholm University.
- Bayliss, A. (2016). Three Tests for Assessing Creativity. *Testing Creativity - Three Tests for Assessing Creativity*. Website. <https://arlonbayliss.com/wp-content/uploads/2016/11/3.Creativity-Tests.pdf>. (accessed Feb 26, 2024)
- Bellera, C.A., Julien, M., & Hanley, J.A. (2010). Normal Approximations to the Distributions of the Wilcoxon Statistics: Accurate to What N? *Graphical Insights. Journal of Statistics Education*, 18(2), 1-17.
- Bsharat, S.M., Myrzakhan, A., & Shen, Z. (2023). Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4. Website. <https://arxiv.org/abs/2312.16171> (accessed Feb 27, 2024)
- Cooper, A. (1999). *The Inmates Are Running the Asylum*. Sams.
- Dwivedi, A.K., Mallawaarachchi, I., & Alvarado, L.A. (2017). Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. *Statistics in medicine*, 36(14), 2187-2205.

- Falk, J., & Young, F. (2022). Supporting Fast Design: The Potential of Hackathons for Co-Creative Systems. *C&C '22: Proceedings of the 14th Conference on Creativity and Cognition*, 515-519.
- Grimes, M., von Krogh, G., Feuerriegel, S., Rink, F., & Gruber, M. (2023). From Scarcity to Abundance: Scholars and Scholarship in an Age of Generative Artificial Intelligence. *Academy of Management Journal*, 66(6), 1617-1624.
- Haase, J., & Hanel, P.H. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. Website. <https://arxiv.org/abs/2303.12003> (accessed Feb 26, 2024)
- Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity*, 34(1), 1-7.
- Hoggenmueller, M., Lupetti, M.L., & van, G.K. (2023). Creative AI for HRI design explorations. In *HRI '23: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (March, 2023), 40-50.
- Hong, M.K., Hakimi, S., Chen, Y.-Y., Toyoda, H., Wu, C., & Klenk, M. (2023). Generative AI for Product Design: Getting the Right Design and the Design Right. In *CHI '23: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (April 23-28, 2023, Hamburg, Germany). ACM, New York, NY, USA, 1-3.
- Karim, M.R., Antar, S.S., & Khan, M.A. (2022). Idea generation using transformer decoder models. In *Proceedings of the 2022 5th International Conference on Algorithms, Computing and Artificial Intelligence*, 1-5.
- Kocaballi, A.B. (2023). Conversational AI-Powered Design: ChatGPT as Designer, User, and Product. Website, <https://arxiv.org/abs/2302.07406> (accessed Feb 27, 2024)
- OpenAI, <https://openai.com/> (accessed Feb 27, 2024)
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. OpenAI. Website. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (accessed Feb 27, 2024)
- Runco, M.A. (2023). AI can only produce artificial creativity. *Journal of Creativity*, 33(72), 1-7.
- Romano, J., Kromrey, J.D., Coraggio, J., & Skowronek, J. (2006). Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys? *Paper presented at the Annual Meeting of the Florida Association of Institutional Research*, February 1-3, Cocoa Beach, Florida.
- Seifert, C.M., & Patalano, A.L. (2001). Opportunism in memory: Preparing for chance encounters. *Current Directions in Psychological Science*, 10(6), 198-201.
- Still, A., & d'Inverno, M. (2019). Can machines be artists? A Deweyan response in theory and practice. *Arts*, 8(1), 1-13.
- Verganti, R., Vendraminelli, L., & Iansiti, M. (2020). Innovation and design in the age of artificial intelligence. *Journal of Product Innovation Management*, 37(3), 212-227.
- Zhou, Y., Muresanu, A.I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2022). Large language models are human-level prompt engineers. Website. <https://arxiv.org/abs/2211.01910> (accessed Feb 28, 2024)
- 布留川英一. (2021). 『BERT/GPT-3/DALL-E 自然言語処理・画像処理・音声処理 人工知能プログラミング実践入門』. 株式会社ボーナデジタル.
- 松本勝. (2021). 『破壊的イノベーションの起こし方：誰でも使えるアイデア創出フレームワーク』. 東洋経済新報社.
- 水本篤, 竹内理 (2008) 「研究論文における効果量の報告のために—基礎的概念と注意点—」『英語教育研究』31, 57-66 頁。
- 水本篤, 竹内理 (2010) 「効果量と検定力分析入門 —統計的検定を正しく使うために—」『外国語教育メディア学会 (LET) 関西支部 メソドロジー研究部会 2010 年度報告論集』47-73 頁, <https://www.mizumot.com/method/mizumoto-takeuchi.pdf> (2024 年 2 月 26 日閲覧)
- 小川亮, 小口裕, 千田彩花 (2024) 「生成 AI の創造性寄与に関する一考察 —チューハイのパッケージデザインを例に—」『マーケティングジャーナル』43, 3, 55-67 頁。
- 大久保街亜, 岡田謙介 (2012). 『伝えるための心理統計』. 勁草書房.
- 山方健士, 湊宣明. (2021). 『リ・デザイン思考法 宇宙開発から生まれた発想ツール』. 実務教育出版.

Human vs. Generative AI: A Comparative Analysis of Creative Idea Generation Capabilities

Takayuki Suzuki*

Abstract:

This study focuses on validating the effectiveness of an ideation method that integrates Generative AI (GPT-4) with the widely used 5WH framework. With the advancements in modern technology, AI has become increasingly important in various creative processes. However, direct comparisons between AI and human involvement have not been thoroughly analyzed. This research compares two different approaches to evaluate the effects of the creative process using GPT-4. The first approach involves humans providing specific instructions (prompts) after which GPT-4 independently generates ideas. The second approach involves ideation solely by humans without AI intervention. Through these approaches, the study aims to compare the impact on four aspects of creativity—originality, diversity, fluency, and elaboration—and to determine which approach excels or falls short in each aspect. The experiment was conducted in a workshop setting with students focusing on ideation for new pickle product development. The results indicated that human-generated ideas scored higher in originality, diversity, and elaboration than those generated by GPT-4. However, no significant difference was observed in fluency between the two. While the study is limited to a specific participant group and the context of new pickle product development, and thus caution is required for generalization to other contexts or conditions, it suggests the potential and limitations of GPT-4 as a complement to human involvement in the ideation process. The findings imply the importance of human creative intervention in generating new perspectives and ideas.

Keywords:

Generative AI, ChatGPT, GPT-4, Idea Generation, Product Development

* Associate Professor, Faculty of Economics and Business Administration, Kyoto University of Advanced Science