SOPHIA U

実測ペルソナを注入したLLMエージェントによる消費者情報探索行動の再現可能性

木岡 良馬　上智大学大学院 応用データサイエンス学位プログラム 修士2年
伴 果純　上智大学大学院 応用データサイエンス学位プログラム 准教授

マーケティングカンファレンス2026 ポスターセッション（2026年10月25日）
データ提供：株式会社ヴァリューズ

1.要旨

概要

消費者が目的を持ってWebを探す場面で、LLMエージェントは渡す情報によって「どこを見るか」を人間に近づけられた。一方、「どれだけ・どう探すか」を近づけるのは、今回の条件では難しかった

	明らかになったこと
どこを見るか	渡す情報を増やすと、**人間との行き先の違い（JSD）は0.165→0.075と半分以下に縮んだ**。ただし人間どうしでも自然に生じる違い（0.027）までは、今回は届かなかった
どれだけ探すか	**1回の探索で10ページ以上見る回**は人間36%に対しエージェント2～3%。今回渡した3種類の情報では、いずれも変化が見られなかった
どう探すか	エージェントは人間の1.4～1.6倍検索に頼る。 人間は**40.6%** の割合でYouTube・X・Amazonなど「いつものサイト」を直接開くが、エージェントは2.1～2.7%（**行動例を渡すと14.6%**）
誰を再現できるか	短く検索中心に探す人は、そのタイプの行動例を渡した条件で、**人間と区別できない水準だった**。ただしこれはエージェント本来のふるまいに近いタイプである

2.背景と目的

背景と先行研究

LLMのシンセティックユーザーとしての実務活用が進み、回答の再現性に関する研究も蓄積されつつある。一方、消費者の行動プロセスに着目した検証は限定的である。

サービスとしての活用事例

✓ **電通の「People Model」**

独自の生活者調査・消費行動データ → 特化型AIペルソナ（1対1やグループ形式） → インタビューによる反応を獲得

電通: https://www.dentsu.co.jp/news/release/2026/0406-011026.html

仕組み	長年蓄積した生活者調査・消費行動データをもとに、約750パターンの多様な生活者像を「特化型AIペルソナ」として再現している
サービス	1対1やグループ形式でAIペルソナにインタビューし、商品や広告に対するターゲット層の反応を素早く把握して、インサイト発見から施策立案までを高速に進める。

先行研究で明らかになっていること

①回答の模倣

本人へのインタビューをもとにしたエージェントは、アンケートの答えを本人とほぼ同じ水準で再現できた。
Park et al.（2024）。本人が2週間後に答え直したときの一致を100%とすると約83%

②個人の再現

一方、年齢・性別などの属性だけでは、一人ひとりの違いを再現するには十分ではない。
属性だけ渡すと約74%（Park et al. 2024）。属性で説明できる人の判断の違いは1割未満（Hu & Collier 2024）

③行動の再現

Lu et al.（2026）は、実世界のオンライン顧客行動ログに基づいてLLMをファインチューニングし、個人のWeb操作行動を再現・予測する枠組みを提示した。ただし対象は1つのECサイト内で、サイトをまたいで探す行動の検証は見当たらない。

※ Z. Wang, Y. Lu, W. Li, et al. OPeRA: A Dataset of Observation, Persona, Rationale, and Action for Evaluating LLMs on Human Online Shopping Behavior Simulation. ACL 2026（arXiv:2506.05606）.
※ J. S. Park, C. Q. Zou, J. Kamphorst, et al. LLM Agents Grounded in Self-Reports Enable General-Purpose Simulation of Individuals. arXiv:2411.10109, 2024（v3, 2026）.
※ Y. Lu, J. Huang, Y. Han, et al. Can LLM Agents Simulate Multi-Turn Human Behavior? Evidence from Real Online Customer Behavior Data. ACL 2026（arXiv:2503.20749）.
※ T. Hu and N. Collier. Quantifying the Persona Effect in LLM Simulations. ACL 2024（arXiv:2402.10811）.

2.背景と目的

研究の目的と問い

本研究は、消費者が目的を持ってWebを探す場面に絞り、エージェントが「どこまで再現できるか」を切り分ける

目的ありの探索行動について

定義	✓ **目的あり探索**＝自分から検索を始め、特定の情報を狙って探す行動 (ログからの取り出し方はスライド7)
実務	✓ シンセティックユーザーで代わりに試したいのは、消費者が必要を感じて商品やサービスを調べ、比べる場面である。広告・SEO・比較サイトなどの施策が効くのもこの場面
比べ方	✓ エージェントは目的（課題）を与えないと動かない。人間側も「目的を持って探し始めた行動」と比べないと、条件の違うものどうしを比べることになる
独自性	✓ ①同一人物の12か月の実ログと意識データ ②1サイト内でなく開放Webをサイト横断 ③日本の消費者

本研究における3つのRQ

RQ1(a)：
目的あり探索全体として、どの行動の側面を再現できるか

RQ1(b)：
消費者のタイプ別に見ると、どのタイプを再現できるか

RQ2：
渡す情報（デモグラ・価値観・行動例）によって、何が再現できるようになるか

3.方法

今回提供いただいたデータについて

同じ1,000人の行動と意識がひもづくヴァリューズ社のデータにより、エージェントを「本人」と比べられる

データ提供元の詳細

提供元	**株式会社ヴァリューズ** 国内250万人規模のWeb行動ログをもとにインターネット行動ログ分析とマーケティング支援を手がける企業である。
どういったデータなのか	Dockpit Perscope サービスとして、同意を得たモニターのWeb行動ログを分析するリサーチエンジン**Dockpit**、行動ログとアンケートを組み合わせる**Perscope**を提供している。 今回はこのDockpitの行動ログとPerscopeの意識データを同一個人で対応づけた1,000人・12か月（2025年6月～2026年5月、約780万回のサイト訪問）のデータを提供いただいた。

データの詳細(変数)

データ	主な変数（★は本研究で使ったもの）
行動ログ（Dockpit）1行＝1回のサイト訪問	★訪問日時（1時間単位）、★サイト（ドメイン・URL）、ページタイトル、サイトのカテゴリ（24分類・55小分類）、★閲覧ページ数（PV）、滞在時間、★流入元（自然検索・外部サイト・直接など）、★検索語（取れる場合のみ）
属性	★年代・★性別・都道府県・★職業・★世帯年収・個人年収・★婚姻・★子供の有無（★の6項目をデモグラとしてエージェントに渡した）
意識（Perscope）	★価値観20項目（例：「自分らしさを大切にしたい」「社会的に高い地位を得たい」）、購入時の重視点、消費行動タイプ、信頼している情報媒体、購買の考え方7項目（例：「高価なものは事前によく調べる」）、メディア接触時間、アプリ利用時間（X・Instagram・YouTubeなど）、興味関心

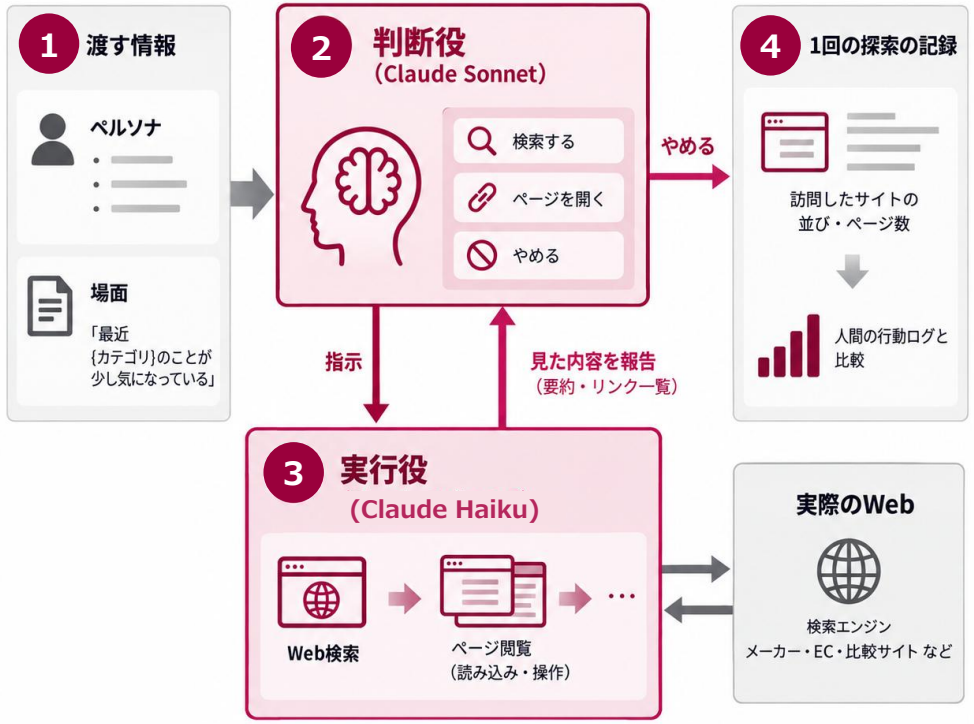
※ 株式会社ヴァリューズについて https://www.valuesccg.com/

3.方法

研究設計

渡す情報をあらかじめ設定し、エージェントに実際のWebで探させ、その記録を実在の消費者の行動ログと同じ物差しで比べる設計

実験環境の設計図



- 結果②では、本実験とは別に3タイプ別の追加実験を行った
- 価値観と行動例を分けて渡す追加実験（各100回）も行った

それぞれの役割と詳細

1. プロンプトとして、情報なし／＋デモグラ／＋価値観・**行動例**の3つの条件でのペルソナを付与し、各100回の実験を行った。場面として、目的を持って探し始める場面を与える。**「最近{カテゴリ}のことが少し気になっている」**消費者として探す。車・旅行・不動産・保険・スキンケアの5カテゴリ*

2. プロンプトによるペルソナ設計のもと**判断役**としてこの消費者を再現するための**行動を選択する**。(検索する、ページを開く、**知っているサイト**に直行する、やめる)

3. 判断役（Claude Sonnet）が次の行動を選び、**実行役**がWeb検索・ページの読み込み・ブラウザ操作のツール（WebSearch・WebFetch・Playwright）で実際に**検索・閲覧する**。やめどきも判断役に任せ、回数の上限は見せない

4. 人間とエージェントを、同じ単位（サイト訪問）・同じサイト分類で比較を行うため蓄積

※ * 5カテゴリは、Googleがヴァリューズの検索データ（180万人分）を分析した消費者調査「バタフライ・サーキット」の対象カテゴリ。消費者が「さぐる」と「かためる」を行き来する一本道ではない情報探索が示されたため、探索行動を実行させる環境として採用した（Think with Google、2020年1月 https://business.google.com/jp/think/measurement/butterflycircuit1/ ）

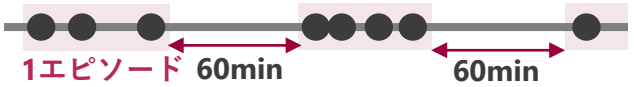
3.方法

分析手法

閲覧記録を60分の空白で1回の探索に区切り、目的を持った探索だけを取り出して、人間とエージェントを5つの側面で、全体とタイプ別に比べる

エピソードの作り方

■ 1人の閲覧記録例



● ＝URL遷移のポイント

✓ 1人の閲覧記録を、**60分以上あいたところで区切り**、一続きの閲覧を「1回の探索(**エピソード**)」とする。

✓ 全体で**約132万回**のエピソード

取り出し方

すべてのエピソード
1,320,361回

検索から始まる
186,666回

検索語が分かる
50,276回

サービス名の検索 -7,088回

ポイ活 -5,015回

同じ語の繰り返し -4,895回

目的あり探索
33,278回・898人

比べ方とタイプ分け

目的あり探索
33,278回・898人

①5つの物差しで人間と比べる

どこを見るか	行き先の構成
どれだけ探すか	探索量・探索の多様性
どう探すか	検索への依存・入り方

②判定
人間から同じ件数を抜き出したときの偶然の幅（95%）に入れば「**区別できない**」

③タイプ分け（→ 結果②）

誰を再現できるか

※除外の理由：サービス名の検索＝サイトへの移動手段、ポイ活＝報酬のための作業（協力者がアンケートモニターのため多い）、同じ語の繰り返し＝習慣的な定型作業

4.結果

結果① 全体の再現

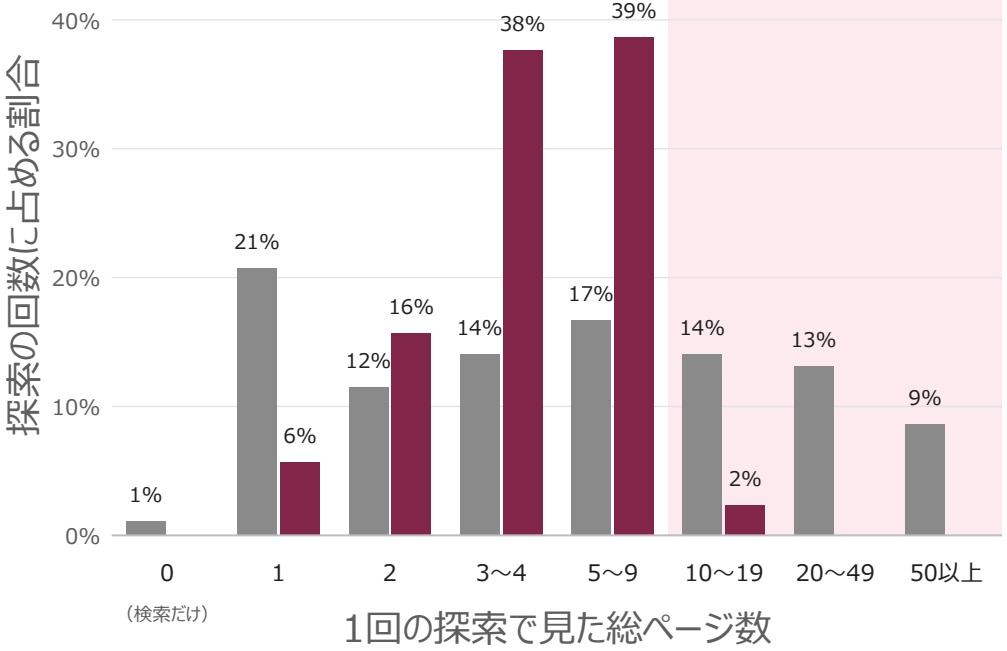
エージェントの探索の長さは、中央値では人間に近かった。一方、長く探す行動と、いつものサイトを直接開く行動はほとんど見られず、代わりに検索に頼る傾向があった

人間とエージェントの探索の長さの比較

人間（目的あり探索 33,278回）
中央値: 5 平均: 17.5

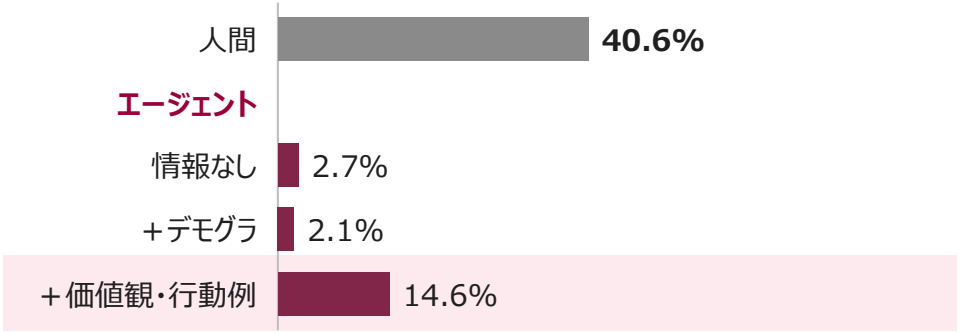
エージェント（主な3条件 計300回）
中央値: 4 平均: 4.3

長く探す回（10ページ以上）
人間 36%
エージェント 2%

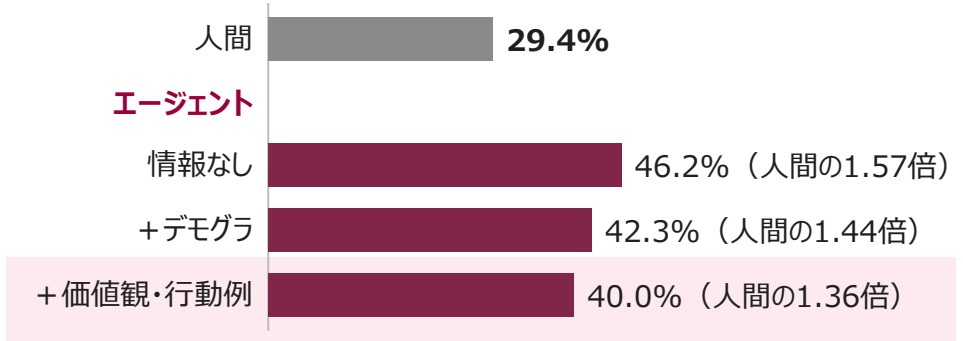


探索行動の違い

①いつものサイト(SNSなど)を直接開く（2つ目以降）



②検索に頼る（サイト訪問に占める検索の割合）



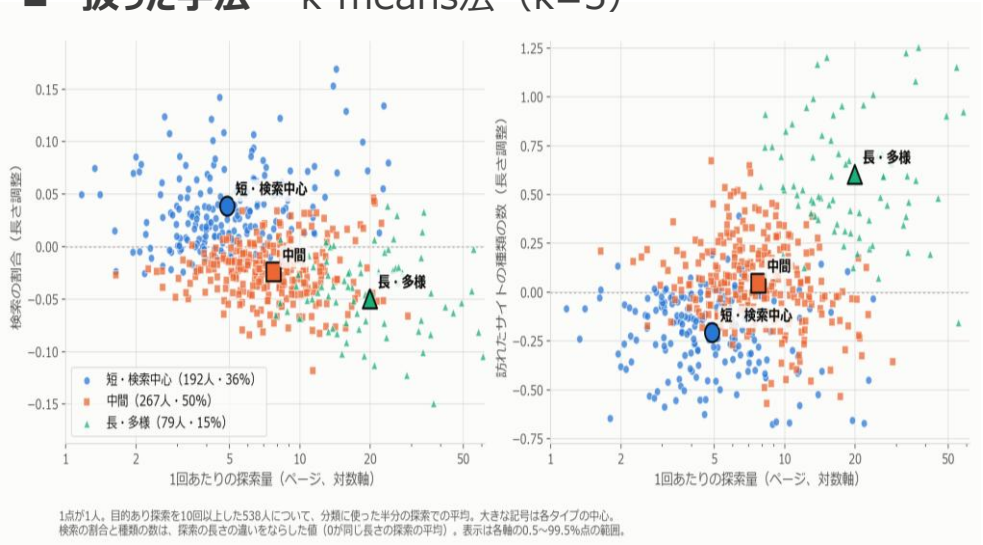
4.結果

結果② 消費者タイプ別の再現

再現の度合いは消費者のタイプで分かれた。「短く検索中心」の人は3つの側面とも人間側の偶然の幅に収まり区別できなかった一方、「長く多様に探す」人は今回は3つとも幅の外であった

人間側の3タイプ(クラスタリング)

■ **扱った手法**　k-means法（k=3）

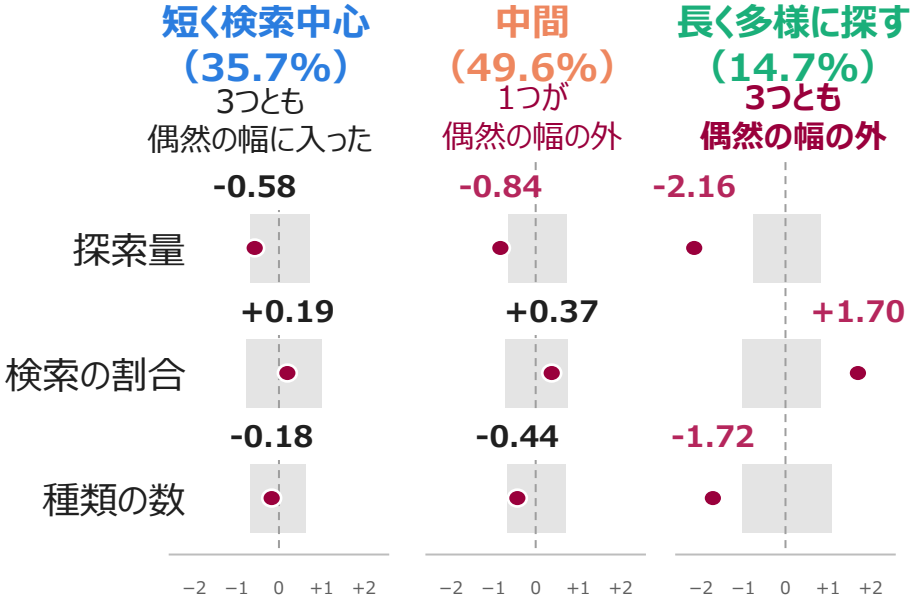


■ **対象**
目的あり探索を10回以上した538人(目的あり探索の95.4%をカバー)

■ **変数**
①探索量（1回の総ページ数の対数）、②検索の割合、③訪れたサイトの種類の数。(②③は、同じ訪問数の回の平均との差)

タイプ別のエージェントのずれ

追加実験：各タイプから8人を選び、デモグラと本人の行動例3件を渡して5カテゴリずつ探索させた（各タイプ40回、計120回）



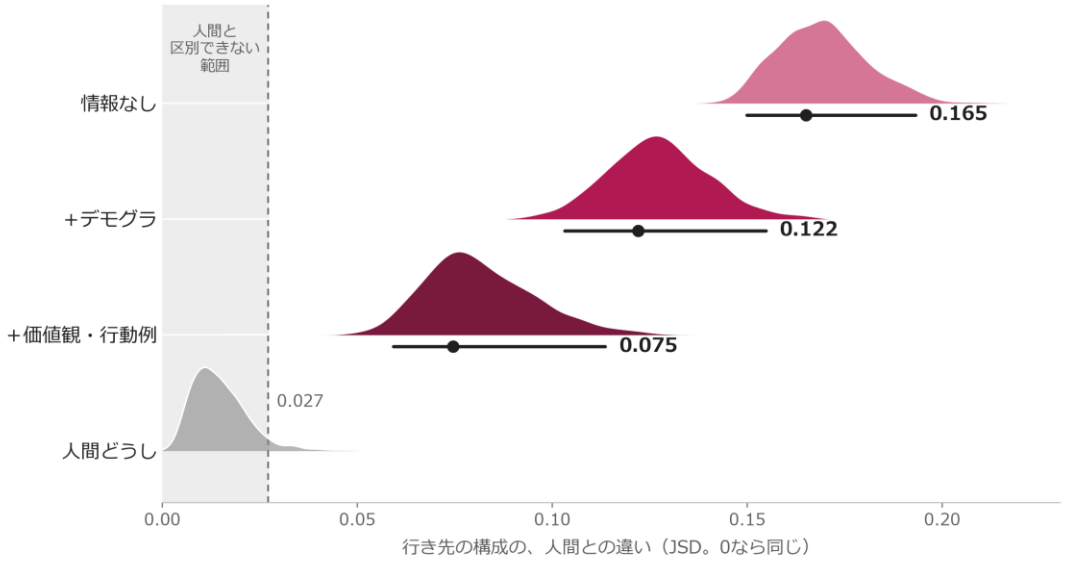
● ＝エージェント（各タイプ8人の行動例を渡した40回の平均）。
点線＝人間の平均
灰色の帯＝人間側の偶然の幅
（同じタイプの8人×5回を抜き出したときの平均の95%範囲）

4.結果

結果③ 渡す情報の効果

情報を足すと「どこを見るか」は人間に近づいた。
「どれだけ探すか」は短い探索が近づいたが、長く探す行動は今回の情報では増えなかった

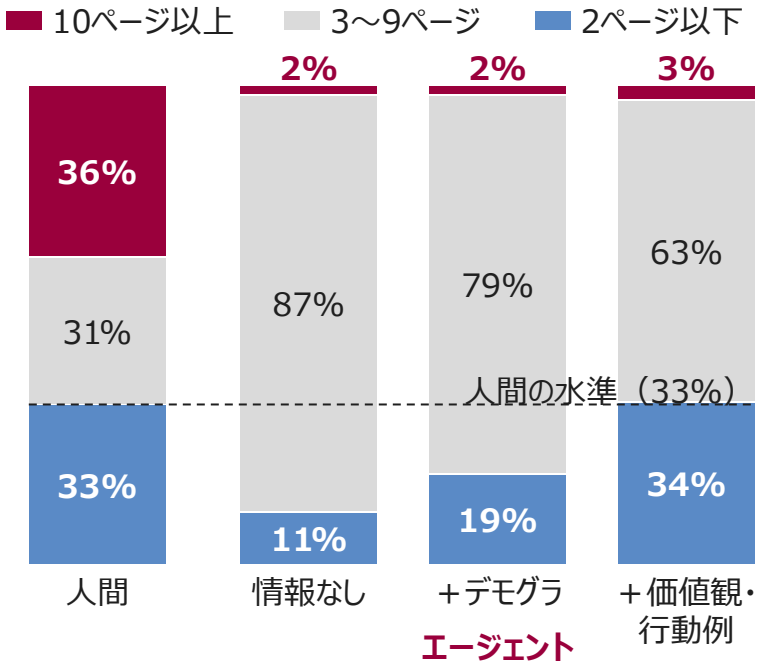
人間とエージェントの行き先の違い



点と線＝推定値と95%区間（各条件100回）。グレーの山＝人間どうしで同じ件数を抜き出して比べたときに自然に出る違い。破線はその95%点

✓ **行き先とは**
どの種類のサイトに何割行ったかの構成を表している
サイト例：検索、メーカー公式、EC,SNS,ニュースポータルなど
結果①でも示した通り、今回のエージェントは探索の行き先として検索に偏っていたことがわかった

人間とエージェントの探索の長さの内訳



1回の探索で見た総ページ数の内訳。
人間＝目的あり探索33,278回、エージェント＝各条件100回

考察

長く探す行動とタイプによる再現の度合いの差は、渡す情報ではなく、今回用いたClaudeのふるまいで説明できると考えられる。課題には状況を、価値観には使い道を与える必要がある

	なぜ長く探さないのか	なぜ行動例のほうが効いたのか
見立て	課題は「最近{カテゴリ}のことが少し気になっている」の1文だけで、目的として弱く、目的だけが独立していた。動機がどこから生まれたのかという状況（シチュエーション）がなかったことが一因と考えられる	行動例は、どの場面で何を使えばよいかが分かり、検索語にもそのまま反映できる。一方、価値観は検索行動のどこに組み込むかの手がかりがなく、検索語にも反映されなかった。本来は、価値観を次の検索の判断に使うと想定していた
裏付け	✓ エージェントがやめた理由は、9割以上が「十分に分かった」「これ以上得るものが少ない」（エージェントの申告）→ 区切りをつけて終えている ✓ 「脱線」でやめたのは4～6%だけ。人間は、目的あり探索でも約3分の1が最後に別の種類のサイトへ移って終わる（ログからの推定）	✓ 渡した情報は検索語に書き込まれる（デモグラでは最初の検索語の4～6割に年代・性別。人間は1%未満） ✓ 価値観では違いは出たが、向きが本人の行動と合わなかった ✓ 行動例も、見せた例を写すだけの場合の一致度には届かなかった
今後へ	目的だけでなく、探し始めた状況を渡す	価値観を、次に何を検索するかの判断に使える形で渡す

6.今後に向けて

今後取り組む予定のアプローチ

探し始めた状況と、価値観の使い道を渡し方に加え、長く探す行動まで再現できるかを確かめる

	①探し始めた状況を渡す	②価値観に使い道を与える
取り組むこと	目的だけでなく、動機がどこから生まれたのかという状況（シチュエーション）を課題に加える	価値観を、次に何を検索するかの判断に使える形で渡す
やり方（案）	✓ 本人の行動ログから、探し始める直前に見ていたもの（SNS・ニュース・広告など）を取り出し、状況として渡す ✓ 人間側も、同じ状況から始まった探索を選んで比べる	✓ 価値観や購買の考え方が、次の検索にどう効くかの例を添える（例：高価なものは事前に調べる → 比較を重ねる） ✓ 行動に近い項目（メディア接触・アプリ利用時間）から使う
確かめること	長く探す回（10ページ以上）と、いつものサイトを開く行動が人間に近づくか	価値観を渡したときの違いが、本人の実際の行動と同じ向きになるか