

# チャットボットの役割分離が 信頼性に与える影響

守屋 彰二<sup>1,2,a)</sup> 大萩 雅也<sup>1</sup> 山崎 天<sup>1</sup> 高山 隼矢<sup>1</sup> 吉川 克正<sup>1</sup> 樋口 啓太<sup>1</sup>

**概要：**オンライン広告ではバナー広告が主流であるが、静的かつ一方であるため、ユーザの文脈や嗜好に応じた推薦には限界がある。生成 AI を活用した対話型の広告は双方向性やパーソナライズを可能にする一方、アシスタント発話に宣伝を挿入する手法は、助言と広告の境界を曖昧にし、信頼を損ねうる。本研究では、非商業的な助言を担うアシスタントと宣伝を担う広告エージェントを分離する役割分離設計 *AdBot* を提案する。役割の分離は、発話者ラベルやインターフェース上の明確な区別表示によって強化する。我々は、バナー広告および発話内広告との比較を通じて、2つの実験で *AdBot* を評価した。製品推薦タスクの予備実験では、対話型手法においてバナー広告よりもユーザの嗜好に適合した推薦が可能であることを示した。さらに、推薦とカウンセリングを対象としたユーザ実験では、*AdBot* は広告の有用性を維持しつつ、発話内広告に比べてアシスタントへの信頼感を向上させることが示された。これらの結果は、対話システムに広告を統合する際、ユーザの信頼を損なうことなく効果を確保する実践的戦略として、役割分離が有効であることを示唆する。

## 1. はじめに

生成 AI の普及に伴い、情報アクセスの主要な窓口は従来の Web インターフェースから、生成 AI を中核とする対話型システムへと拡大している [1, 2]。検索や相談などの文脈において、静的なスクロール型閲覧が適する場合と、アシスタントを介した対話が有効な場合が併存し、両者を使い分けることが一般的になりつつある [1, 3]。この変化を踏まえ、チャットボット時代にふさわしい広告提示の設計原則を再検討する必要がある。

オンライン広告の主流はバナー広告であり [4, 5]、Web サイト、モバイルアプリケーション、ソーシャルメディアなどに横断的に配信できる。一方で提示は一方的であり、ユーザの直近の意図や文脈への適応は限定的である [6, 7]。この限界に対応する方策として、生成 AI 駆動の対話型システムに広告を統合する試みが進められている [8, 9]。例えば Bing Chat <sup>\*1</sup> のように、チャットボットの応答内に広告内容を組み込む方式が検討されている [8, 10]。この方式は、ユーザの発話や意図に応じて説得の焦点をリアルタイムに調整でき、対話を通じて広告内容に関する質疑応答を可能にする。ユーザとのインタラクティブなやり取りを

介して、ユーザの関心への適合や商品理解の深化が期待される [11]。

しかし、アシスタントの応答そのものに広告を挿入すると、アシスタントへの信頼を損なう懸念がある [8, 12]。ユーザは中立的かつ誠実な助言を期待するが、第三者の商業的利益に基づく推薦が回答に混入すると、不信につながりうる。例えば、「おすすめの旅行先は？」というユーザの問いに対し、「バルセロナがおすすめです。広告: バルセロナを訪れる際はぜひストア X へ」と続く応答では、前者は助言、後者が広告であることを明示的に区別しても、対話の連続性ゆえに助言が広告意図に影響されたものと受け取られかねない。この影響は、誠実さが特に重視されるメンタルヘルス支援やキャリア相談といった文脈で顕著であると考えられる。

本研究では、この信頼低下の問題に対処するため、役割分離に基づく二エージェント構成を提案する。システムは、広告から独立した助言を与える *MainBot* と、広告の提示・説明に特化した *AdBot* から構成される (図 1)。情報源と発話責任の帰属を分離し、発話者ラベル、メタデータ、インターフェース上の明示な区別表示によって、*MainBot* と *AdBot* の意図を可視化する。これにより、広告に伴う不信感を *AdBot* 側に限定し、*MainBot* の助言に対する信頼を維持することを狙う。さらに、バナー広告に対する主要な利点である対話的探索を保持するため、ユーザは必要に応

<sup>1</sup> SB Intuitions

<sup>2</sup> 東北大学

<sup>a)</sup> shoji.moriya@sbintuitions.co.jp

<sup>\*1</sup> <https://www.bing.com>



図 1 チャットボット・インターフェースにおける 3 種類の広告提示形式 (製品推薦シナリオ).

じて対話相手を *AdBot* に切り替え、広告に関する追加質問や説明要求を行える設計とする。

実証的検証として、我々は広告の有用性が異なる 2 つの対話シナリオ (製品推薦とカウンセリング) を設定し、ユーザスタディを実施した。比較条件は、バナー広告、発話内広告、および提案手法である *AdBot* である。評価は、広告とユーザ嗜好の一致度、アシスタントへの信頼、有用性、ユーザ体験などの複数指標で行った。予備実験では、対話に基づく二方式 (発話内広告と *AdBot*) が、バナー広告よりユーザの嗜好に合致したアイテムを提示できることを示した。さらにユーザスタディでは、*AdBot* は発話内広告に比べてアシスタントへの信頼を向上させつつ、広告の有用性に有意差は認められなかった。

本研究の貢献は次のとおりである。

- 助言と広告を分離する二エージェント設計 (*Main-Bot/AdBot*) を提案した。
- シミュレータ評価では、対話型の方式がバナー広告よりもユーザの嗜好に合致したアイテムを提示できることを示し、比較ユーザスタディでは、役割分離により発話内広告より高い信頼を保持しつつ、有用性は同程度に保たれることを確認した。
- これらの知見に基づき、対話システムへの広告統合に関する設計指針を提示した。

## 2. 関連研究

### 2.1 LLM ベースの AI アシスタント

大規模言語モデル (LLM) の登場により、対話型インターフェースは日常的な知的タスクのアシスタントとして広

く利用されている [13]。応用領域は情報検索 [13,14]、要約 [15]、カウンセリング [16] などへと拡大し、人間と AI のインタラクションにおける期待や依存、検証行動の変容が検討されている。実運用の文脈では、タスク遂行時間の短縮や成果物の品質向上が効果が報告され、実用効果の高まりが示唆されている [17]。また、学習指向型検索に関するタスクベース調査では、人々は生成 AI チャットを迅速な「要約・足場づくり」と位置づけ、従来の Web 検索を置き換えるのではなく補完的に活用していた [2]。

一方で、AI アシスタントへの信頼は一様ではない。ChatGPT に対する信用は、ゲートキーピング認識、情報の網羅性への懸念、インタラクション様式などの要因に左右される [18]。他方で、健康情報探索など一部のタスクでは、ChatGPT の応答がユーザの主観的信頼基準を上回る場合も報告されている [19]。総じて、アシスタント型利用において妥当な信頼を持続させるためのインターフェース設計が求められる。

### 2.2 広告提示形式

バナー広告は、Web やモバイルアプリ、ソーシャルネットワークなどに広く用いられてきた従来型のオンライン広告である [5]。広域配信が容易である一方、ユーザが広告領域を無意識に見落とす「バナーブラインドネス」に陥りやすいことが知られており [20,21]、この効果はアイトラッキングやタスク実験で確認されている。結果として、関与の低下やクリック率の伸び悩みといった実務上の課題に直結する [4,21]。

近年、対話型 AI の登場に伴い、チャットボットの応答内

に広告を組み込む試みが登場している [8,9]. Bing Chat <sup>\*1</sup> に代表されるこの方式は、発話内広告あるいは生成型ネイティブ広告と呼ばれ、ユーザの発話や意図に応じて双方向にメッセージを提示できる。対話を通じた関心の探索により、興味への適合性を高め、理解の促進が期待される [22,23].

一方で、課題も指摘されている。生成応答と広告との境界が曖昧になると、ユーザが当該メッセージを広告と認識できない倫理的懸念が生じる [24,25]. これを緩和するため、開示文言や配置の最適化、明確な広告ラベル付与が提案されてきたが、こうした措置はスポンサーや媒体に対する評価を下げる可能性も指摘される [12]. さらに、チャットボットを情報提供者やアシスタントとして信頼するユーザに対し、応答に利害関係を伴う広告推薦が混入すると、中立的助言への期待を損ないうる [8,12].

このため、効果的な広告宣伝を実現しつつ、アシスタントへの信頼を維持する設計が求められる。本研究は、この課題に応えるものとして、広告に特化した専用エージェント *AdBot* を導入する。

### 3. AdBot: 役割分離型二エージェント方式

#### 3.1 設計コンセプト

チャットボットに広告を統合する基本設計として、(i) 対話と独立した領域にバナー広告を配置する方式、(ii) チャットボットの応答内に推薦を組み込む発話内広告方式、の二つが考えられる。バナー広告は、標準化されたスロットに対話とは分離して表示されるため、チャット体験を阻害しにくく、制作や運用コストも抑えやすい。一方で、提示が一方向であるため、ユーザの意図に合わせた柔軟な調整や、広告内容の深掘りには限界がある。これに対し、発話内広告は双方向のやり取りを通じて、ユーザのニーズに即した提示が可能である。しかし、商業的利害を伴う情報をアシスタントの応答に直接混在させ、アシスタントと広告主を同一の話者として提示することは、応答の独立性やシステム全体への信頼に関する懸念を招く。

以上の課題に対処するため、我々は役割分離に基づく二エージェント方式を提案する。広告的利害から独立して助言を行う *MainBot* と、広告の提示・説明を担う *AdBot* を組み合わせる。これにより、バナー広告の一方向性という制約を回避しつつ、文脈に応じた対話的な深掘りを可能にする。ユーザは必要に応じて *AdBot* と対話を開始し、関連性の確認や詳細説明を求めることができる。さらに、発話レベルで情報源を分離することで、広告に伴う不信感を *AdBot* 側に局所化し、*MainBot* への信頼を維持できる。

自動化や対話エージェントへの信頼を多角的に概念化した先行研究 [26] に基づき、本稿では信頼のうち、特に「助言を行うアシスタント (*MainBot*) が商業的インセンティブに不当に影響されず、公平な助言を提供しているか」とい

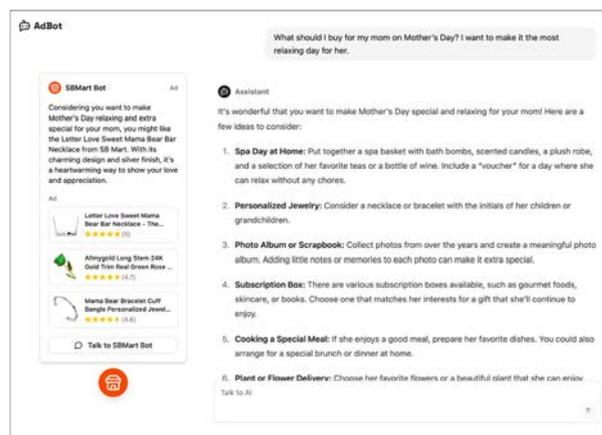


図 2 *MainBot* モード：*MainBot* の発話後に左側にポップアップが表示される。ポップアップは開閉やドラッグが可能である。*AdBot* は「SBMart Bot」として表示している。

Whatever you choose, make sure it aligns with her interests and personal preferences. It's the thought and effort you put into making it her special day that will be most appreciated.

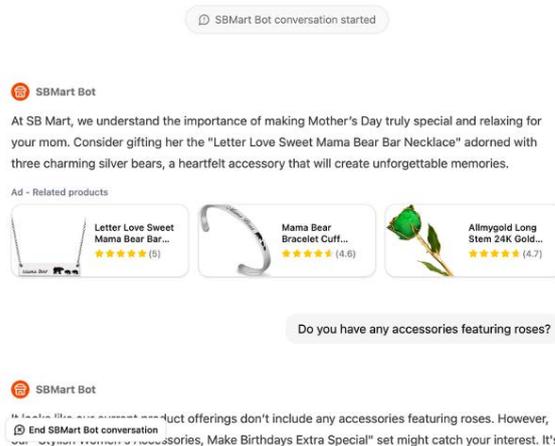


図 3 *AdBot* モード：広告領域を明確に認識できるよう分離された対話。*AdBot* は「SBMart Bot」として表示している。

うユーザの信念を指す。基盤となる LLM と検索パイプラインを共通化することで、条件間における知能の知覚と総合的な有用性を一定に保ち、代わりにエージェント間で商業的意図がどのように帰属・分離されるかを操作する。

#### 3.2 設計デザイン

提案手法は、中立的なアシスタント (*MainBot*) と広告エージェント (*AdBot*) から構成される二エージェント方式である。*MainBot* はユーザ発話に対して広告を含まない応答を生成し、*AdBot* は広告の提示・説明を担当する。*MainBot* がユーザの目標との整合のみに最適化されるのに対し、*AdBot* はユーザの質問と広告主の利害の双方への整合という二重の目的を担う点が特徴である。

図 2 に示すように、ユーザはまず *MainBot* との対話から開始する。この間、*AdBot* は関連するアイテムが見つかった場合に、メイン対話に隣接するポップアップで短い

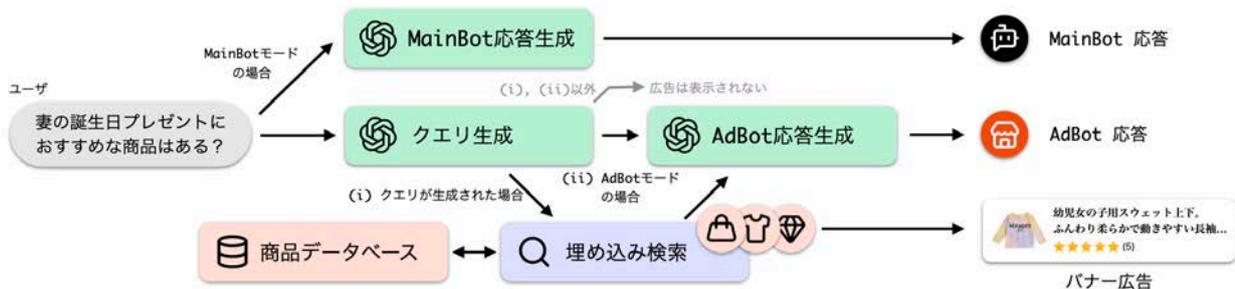


図 4 AdBot システムの概要図。

宣伝メッセージを表示する。これを *MainBot* モードと呼ぶ。該当アイテムがなければ広告は表示されず、*MainBot* の応答のみが提示される。

もう一つのモードが *AdBot* モード (図 3) である。ユーザは明示的に *AdBot* を呼び出し、広告対象に関するフォローアップ質問やアイテム探索などのやり取りを行える。ポップアップ内のボタンで *AdBot* モードへ切り替え、終了ボタンを押すことで任意のタイミングで *MainBot* モードに戻ることができる。戻際には、*AdBot* との対話履歴を保持し、文脈的な一貫性を確保する。この二モード設計により、ユーザは広告への関与度合いを自ら制御できる。

両エージェントには、役割に応じた明確な指示を与える。*MainBot* には、中立的かつ広告を含まない応答を生成し、ユーザ支援に専念させる。一方、*AdBot* には、関連するアイテムが見つかった場合に宣伝コンテンツを応答に組み込み、質問応答とアイテム探索を支援するよう指示する。

図 4 に応答生成の流れを示す。ユーザが発話を送信すると、システムはまずブラウザから送られるアクティブモードを確認する。*MainBot* モードでは、*MainBot* が中立的な応答を生成すると同時に、クエリ生成器が発話をアイテム関連クエリへ変換する。適切なクエリが得られれば、アイテムデータベースに対して埋め込み検索を実行し、上位アイテムを取得してポップアップで提示する。クエリ生成と関連アイテム取得の双方が成功した場合のみ、*AdBot* が推薦を提示し、いずれかが不成立なら *MainBot* の応答のみを返す。*AdBot* モードでは、*MainBot* は介在せず、*AdBot* が応答を生成する。同様にクエリ生成と埋め込み検索を行い、取得アイテムを *AdBot* の応答に統合し、併せてバナー領域にも表示する。クエリが生成されない場合は、*AdBot* は広告を含まない応答を返す。

## 4. 実験

### 4.1 リサーチクエスト

本実験では、提案する二エージェント方式を既存形式 (バナー広告と発話内広告) と比較し、「発話内広告と比較して、*AdBot* はアシスタントへの信頼度をより高く維持

できるか?」というリサーチクエストに答えることを目的とする。ここで「アシスタント」とは、広告利益から独立して設計された助言提供エージェント (*MainBot*) を指す。さらに、対話型形式がバナー広告と比較した推奨アイテムの嗜好適合性にどう関わるかや、*AdBot* と発話内広告間で広告有用性がどう異なるかを検証し、これらの観察結果から設計上の示唆を提示する。

### 4.2 実験設定

#### 4.2.1 タスクシナリオ

本研究では、システムの信頼性が特に重要となる 2 文脈を対象とした。(i) 広告内容がタスク達成に寄与しやすい製品推薦シナリオと、(ii) 広告がタスク達成に直接的な有用性を持たないカウンセリングシナリオである。タスク反復による学習効果を緩和するため、各文脈につき 3 種類の異なるシナリオを準備した。

##### 4.2.1.1 製品推薦

参加者は、親族・知人への贈り物の選定についてシステムに相談した。目標は、何を贈るかについて具体案に到達することである。

- 来週に誕生日を迎える親戚へのプレゼントを探している。今年はおしゃれな衣類を贈りたい。相手の好みと予算を伝えたくて、どのアイテムがよいかシステムに相談する予定である。
- 親しい友人が結婚することになった。式には出席しないが、お祝いとして衣類を贈りたい。相手との関係性・好み・予算を伝えたくて、結婚祝いとしてふさわしい品についてシステムに相談する予定である。
- 近々、親戚の家を訪ねる予定がある。そこには小学校低学年の子どもがいるため、ちょっとした贈り物を用意したい。喜ばれそうなおもちゃを選びたく、どのアイテムがよいかシステムに相談する予定である。

##### 4.2.1.2 カウンセリング

参加者は、不安や悩みの軽減を目的にシステムに相談した。最終的な目標は、懸念を解消することである。達成度は、参加者自身の主観で評価された。

- 仕事を始めてから、学生時代の友人との関係にストレスを感じるようになった。最近はその友人の言動に傲慢さをより強く感じ、自分が見下されているように思うことが増えた。関係を続けるべきか、一定の距離を置くべきか迷っており、気持ちを整理するためにシステムに相談する予定である。
- 仕事や日常生活に大きな問題はないものの、「このままで本当に良いのか」と将来に対する漠然とした不安がある。何かを変えたいが具体的に何をすればよいか分からず、このもやもやが続いている。この心境をシステムに打ち明け、対話を通じて考えを整理する予定である。
- これまではバスケットボールやフットサル、旅行などで休日を楽しんでいたが、最近はずっと忙しく、週末は疲れで寝て過ごすことが多い。SNSで友人たちが楽しむ様子を見ると羨ましく、ときには嫉妬を覚えることもある。仕事自体は楽しいが、うまくリフレッシュできていないと感じている。何をすべきか分からず、明確な解決策がないことも承知のうえで、話を聞いてほしいと考えている。

#### 4.2.2 アイテムデータベースの構築

広告用アイテムデータベースは、公開データとLLMによる合成データを組み合わせて構築した。公開データとしては、“Amazon Reviews’23 dataset” [27] のうち “Amazon Fashion” および “Toys and Games” から 2,000 件ずつ抽出し、GPT-4o で日本語に翻訳した。

合成データは、各実験シナリオごとに 100 件を GPT-4o で生成した。手順は以下の通りである。まず、各シナリオに対して 10 種類の広告ジャンルを生成した。例えば、カウンセリング関連のシナリオでは、「自己啓発書」や「オンラインサロン」等が広告ジャンルとして生成される。次に、各ジャンルにつき 10 件の詳細なアイテム情報を作成した。さらに、合成アイテムに対応する広告画像も GPT-4o で生成した。合成例は、表 1 に示す。

生成したデータセットは人手でチェックを行い、各シナリオにおける広告ジャンルの適切性や多様性、生成アイテムのジャンルとの関連性を確認した。不適切な出力については、再生成または人手で修正を行い、必要に応じてデータセットから除外した。最終的に、既存データと併せて合計 2,588 件の広告データベースを得た。

#### 4.2.3 埋め込みベース検索

広告推薦は埋め込みベース検索により関連アイテムを特定する。生成されたユーザークエリとアイテムは、共通の意味埋め込み空間で表現する。埋め込みには OpenAI の *text-embedding-3-small*\*2 を用いる。

各アイテムについて、タイトル・説明・レビュー文を連

\*2 <https://platform.openai.com/docs/models/text-embedding-3-small>

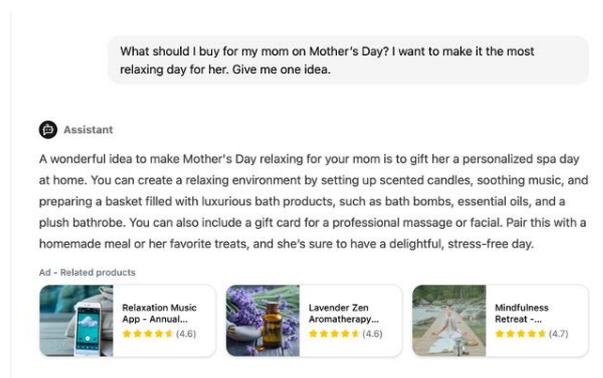


図 5 バナー広告：取得されたアイテムを、アシスタントの応答の下に専用のバナー領域として表示する。

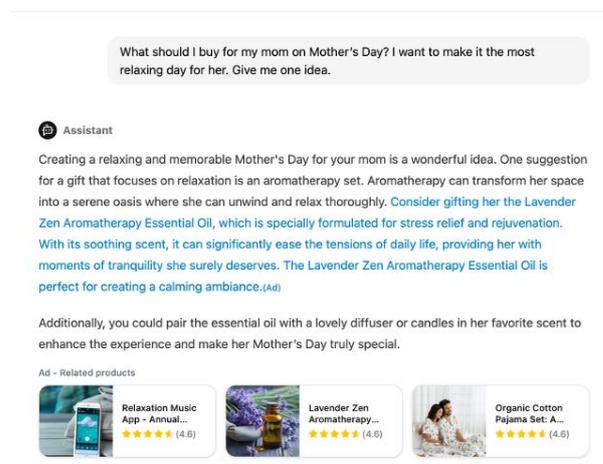


図 6 発話内広告：取得されたアイテムを青色でアシスタントの応答文中に織り込みつつ、メッセージ下にも併せて表示する。

結して埋め込みを計算する。クエリとアイテム埋め込みの類似度は、近傍探索ライブラリ *voyager*\*3 で算出し、上位 3 件を取得して表示コンポーネントへ渡す。

#### 4.2.4 比較手法

*AdBot* を、対話システムおよび Web システムで一般的な 2 つの広告提示形式 (バナー広告/発話内広告) と比較する。いずれの条件でも、広告アイテムの検索には 4.2.3 節と同一の埋め込みベース手法を用い、差分は取得アイテムの提示方法のみである。広告の出現頻度も *AdBot* と同様に、関連アイテムが取得された場合にのみ表示とする。LLM は全条件で GPT-4o\*4 を用いる。

##### 4.2.4.1 バナー広告

図 5 はバナー広告条件の例である。アシスタントの生成テキストは広告を含まず、関連アイテムは応答下部に専用バナー領域に表示される。

##### 4.2.4.2 発話内広告

図 6 は発話内広告条件の例である。広告はアシスタントの応答文中に直接埋め込まれ、検出アイテムは文中に青色で強調表示される。併せて、関連アイテムを応答下部にリ

\*3 <https://github.com/spotify/voyager>

\*4 OpenAI Docs – GPT-4o

表 1 合成した広告アイテムの例. 各広告アイテムのタイトルと説明は, 対応するシナリオおよび広告ジャンルに適合するように設計した.

シナリオ	広告ジャンル	広告名	詳細
製品推薦	メンズファッション	スタイリッシュなスニーカー. カジュアルコーデに合わせやすい.	誕生日ギフトとしても大人気. 快適な履き心地と洗練デザインで, 毎日の普段使いにも最適です.
カウンセリング	マインドフルネスアプリ	Aura: 毎日のストレスや不安を和らげる, パーソナライズド瞑想アプリ.	気分や状態に合わせて最適なセッションを提供し, 人間関係のストレスや価値観の違いによる悩みの軽減をサポートします. 高評価のユーザーレビューでも効果が示されています.

ストとして提示する点はバナー条件と同様である.

### 4.3 予備実験

予備実験では, チャットボットが自然言語で広告を提示することにより, バナー広告形式と比較してユーザの嗜好により適合した商品を提示できるかを検証した. この利点は既存研究でも検証されている [22,23] が, 本研究のシステムにおいても同様の傾向が見られるかを確認した.

#### 4.3.1 予備実験設定

本予備実験では, バナー広告, 発話内広告, *AdBot* を, LLM ベースのユーザシミュレータを用いて評価した. 各シミュレータは, 推薦タスクにおけるターゲットアイテムの属性に整合するペルソナを付与したうえでボットと対話する. 例えば, ターゲットがアウトドア用シャツの場合, 「ハイキングを楽しむ」「機能的な服が好き」などのペルソナが割り当てられる. ボットの推薦にターゲット商品が含まれる場合にセッションは成功と判定される. 各セッションごとにこれを判定し, 全セッション中成功したセッションの割合を「推薦精度」として定義し, 指標として採用した.

実験では, アイテムデータベースから 100 件のターゲットアイテムをサンプリングし, 各ターゲットに対して GPT-4o で 10 種類のペルソナを合成した. ユーザシミュレータ自体も GPT-4o で構築した.

ユーザシミュレータとボットの対話は 2 ターンに固定した. ユーザシミュレータが発話し, ボットがアイテム推薦とともに応答を返す. これを 2 ターン分行った. 対話を 2 ターンに制限したのは, 設定を単純化し, 最小限のやり取りで推薦性能の差を評価するためである.

広告形式による知覚差を模擬するため, ユーザシミュレータに提示する情報を統制した. 人間はバナー画像のような固定的な視覚表示を見落としやすい (バナーブラインドネス) ことが知られており [20,21], このような知覚バイアスはシミュレータでは再現が難しい. 条件間の公平性を保つため, 本実験ではシミュレータにはテキストのみを提示し, 画面上のアイテム画像やキャプションは除外した.

各シミュレータに提供する情報は以下のとおりである:

- (1) バナー広告: 広告情報を含まない一般的な応答
- (2) 発話内広告: 同一ターン内に広告情報が埋め込まれた応答

- (3) *AdBot*: (i) *MainBot* による一般応答 と (ii) *AdBot* による広告関連応答 の分離された応答

シミュレータには発話テキストのみを提示する一方, 評価のためにボット側は画面表示用のアイテム情報も内部で返す. 推薦の一致判定は, このアイテム情報に基づいて行った.

### 4.4 結果

100 件のシミュレーションセッションに基づく評価の結果, 対話ベースの手法はいずれもバナー広告を上回った (バナー広告: 26%, 発話内広告: 38%, *AdBot*: 39%). この優位性は, 対話エージェントが広告情報を明示的に応答に組み込み, ユーザシミュレータからのフォローアップ発話を誘発することでターゲティングが精緻化されるためと考えられる. 一方, バナー方式も一定割合で関連アイテムを取得したが, 精度は対話型条件を下回った.

もっとも, ユーザシミュレータが実際のユーザ行動を完全に再現しているとは限らない. 特に, 実環境ではバナーブラインドネスの影響により, バナー広告との相互作用はさらに低下する可能性が高い. これらを踏まえると, 本システムにおいても対話型推薦が一定の有効性を持つことが示唆される.

## 5. ユーザスタディ

本研究では, 発話内広告と比較して *AdBot* がアシスタントへの信頼をより高く維持できるという仮説を検証するため, ユーザスタディを実施した. 参加者は, 提案手法を含む複数の広告提示形式に対応した対話システムと実際に対話し, 信頼性, 有用性, ユーザ体験を評価した.

実験は所属機関の倫理審査委員会による承認手続きを経て実施された. 参加者は合計 30 名で, 事前調査による内訳は男性 12 名 (40.0%), 女性 18 名 (60.0%) であった. 年齢は 20 代 1 名 (3.3%), 30 代 13 名 (43.3%), 40 代 12 名 (40.0%), 50 代 3 名 (10.0%), 60 歳以上 1 名 (3.3%) であった. チャットボット利用経験については, 日常的に利用している者が 14 名 (46.7%), 時々利用する者が 10 名 (33.3%), 数回程度のみ利用した者が 6 名 (20.0%), ほとんど利用経験のない者はいなかった (0.0%).

## 5.1 手順

- (1) **導入:** 参加者は研究手順と注意事項の説明後、インフォームドコンセントに同意し、年齢、性別、チャットボット利用経験などの属性を回答した。
- (2) **タスク慣熟:** 広告を提示しないシステムと自由に雑談し、インターフェースに慣れてもらった。
- (3) **主要タスク:** 2つのシナリオ × 3つの広告提示形式の計6条件を実施した。各条件で最大20ターンまで対話し、タスク達成と判断した時点で早期に終了することも可能とした。各条件終了後に条件別評価を実施した。
- (4) **事後評価:** 全条件終了後に総合評価を実施した。

シナリオと広告提示形式の順序は参加者間でカウンターバランスを取った。実験で使用した広告データセットは予備実験と同一である（構築手順は4.2.2節を参照）。全形式において基盤モデルとしてGPT-4oを用い、同一データソース・同一手順で広告候補を検索したため、提示頻度は形式間で一定に保たれた。実験時間は、全体を通して約180分であった。

## 5.2 評価指標

各条件終了直後に参加者は条件別評価を、すべての条件終了後に総合評価を行った。

条件別評価では、タスク達成度 (Q1)、広告の顕在性 (Q2)、広告の有用性 (Q3)、広告に対する忌避感 (Q4)、アシスタントの信頼性 (Q5) を五件法リッカート尺度で測定し、ユーザ体験については七件法のUEQ-Sを用いた。Q5では、参加者は非広告アシスタントの信頼性を評価対象とした。すなわち、バナー広告、発話内広告、およびAdBot条件におけるMainBotが評価対象となった。

各シナリオ・各質問ごとに、繰り返し測定に対応したノンパラメトリック検定であるFriedman検定を用いてシステム間比較を行った。その後、3条件間の全ペアに対してWilcoxonの符号付順位検定を行った。検定はいずれも両側検定であり、有意水準は $\alpha = 0.05$ とした。また、各シナリオ・質問における3つのペア比較について、両側 $p$ 値をBenjamini-Hochberg法でFDR補正を適用し、補正後の値を $q_{BH}$ として報告した。

平均差は $A-B$  (Mean diff) として95%信頼区間とともに報告した。信頼区間は参加者ごとの対応差に基づき、3,000回のパーセントイル・ブートストラップにより算出した。さらに、効果量としてペアごとのCohenの $d_z$ を95%ブートストラップ信頼区間とともに提示した。Wilcoxon検定については標準的な同順位/ゼロ差の扱いに従った。

総合評価では、広告形式の順位づけ選好と、各形式の長所・短所に関する自由記述を収集した。自由記述の質的分析は、6.2節に示す。

## 5.3 結果

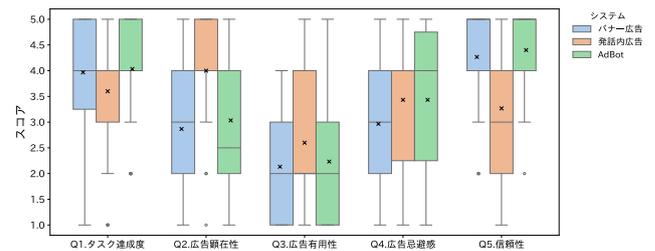


図7 製品推薦シナリオにおける条件別評価結果。各箱（ボックス）は四分位範囲（IQR）を示し、中央の水平線は中央値を表す。ひげは $1.5 \times \text{IQR}$ まで、点は外れ値、×印は平均値を示す。

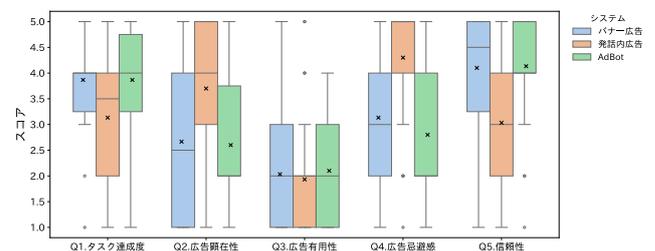


図8 カウンセリングシナリオにおける条件別評価結果。各箱（ボックス）は四分位範囲（IQR）を示し、中央の水平線は中央値を表す。ひげは $1.5 \times \text{IQR}$ まで、点は外れ値、×印は平均値を示す。

各シナリオにおける条件別結果は図7および図8に示す。

信頼性 (Q5) は両シナリオにおいて発話内広告が最も低い値を示した。特にカウンセリングシナリオではこの効果が顕著であり、AdBotが発話内広告を平均で1ポイント以上上回った (5件法,  $d_z = 0.62$ ,  $q_{BH} = 0.006$ )。さらに、発話内広告は他の2形式よりもタスク達成度 (Q1) が有意に低く、広告の顕在性 (Q2) と広告忌避感 (Q4) が有意に高かった。一方、広告の有用性 (Q3) では、いずれのシナリオでも有意なペア差は観測されなかった。

UEQ-Sによるユーザ体験評価では、表2および表3に示すように、実用品質は両シナリオでバナー広告よりも発話内広告が低く、またカウンセリングではAdBotよりも低かった。一方、快楽的品質については両シナリオとも有意差は見られなかった。

総合すると、発話内広告は他形式よりも広告への注意を集めやすい一方で、信頼を損ないやすい傾向が示された。特に、広告がタスク成功に直結しにくいカウンセリングでは、発話内広告における広告忌避感の高さとタスク達成度の低さが顕著であった。

形式ごとの順位づけ選好は、図9に示すとおり条件別結果と整合的であった。参加者のコメントを表4に示す。これらの分析は、6.2節に記す。

表 2 製品推薦シナリオにおけるシステム別 UEQ-S の結果.

システム	実用的品質		快楽的品質		Overall	
	Mean	SD	Mean	SD	Mean	SD
バナー広告	1.400	0.995	-0.125	1.451	0.638	1.102
発話内広告	0.550	1.241	-0.325	1.323	0.113	1.039
AdBot	1.142	1.245	-0.083	1.415	0.529	1.075

表 3 カウンセリングシナリオにおけるシステム別 UEQ-S の結果.

システム	実用的品質		快楽的品質		Overall	
	Mean	SD	Mean	SD	Mean	SD
バナー広告	1.050	1.332	-0.508	1.395	0.271	1.209
発話内広告	0.133	1.436	-0.550	1.436	-0.208	1.248
AdBot	1.008	1.248	-0.242	1.245	0.383	1.019

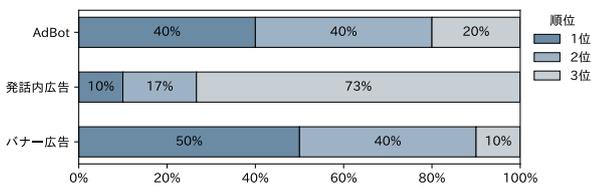


図 9 広告形式の順位づけ選好の割合.

## 6. 議論

### 6.1 リサーチクエストへの回答

本節では、「発話内広告と比較して、AdBot はアシスタントへの信頼度をより高く維持できるか?」というリサーチクエストに沿って結果を解釈する。5.3 節のユーザスタディでは、AdBot は両シナリオで一貫して発話内広告より高い信頼評価を得た。発話内広告は広告への注意を喚起した一方で、知覚される有用性は有意に高まらなかった。特に広告がタスク達成に直接寄与しにくいカウンセリングでは、発話内広告における広告忌避感の上昇とタスク達成度も低下が顕著であった。これらを総合すると、MainBot と AdBot の役割分離は、発話の帰属を明確化により広告に対する抵抗感を局所化し、MainBot の助言の中立性を維持する設計原理として機能することが示唆される。

これに加え、対話ベースの方式では Banner よりもユーザーの嗜好に合致した推奨が表示されることも確認された。

### 6.2 定性分析

数値指標のみでは捉えきれないシステムに対する印象を明らかにするため、参加者から収集した自由記述回答を分析した。総合評価では、参加者に対して各システムの良い点・悪い点をそれぞれ最大 3 項目ずつ記述するよう求めた。本分析では、各システムについて肯定的・否定的なコメント群から重複を避けつつ複数コメントを選定し、原文の意図を変えない範囲で誤字脱字の最小限の修正と匿名化を施した状態で提示する。

結果を表 4 に示す。参加者のコメントは大きく三つのテーマに集約された。第一にレイアウトや可視性とそれに伴う広告の顕在性、第二に文脈の適合性、第三に助言と広告の境界の明確さである。各形式に関するコメントの要約は以下のとおりである。

#### 6.2.1 バナー広告

本文と視覚的に分離され、侵入度が低いため、対話の流れを妨害しにくい点が好意的に評価された。一方で、情報のやり取りが単方向に限定されるため、ユーザ意図を掘り下げて適合性を高めることが難しいという指摘もあった。

#### 6.2.2 発話内広告

対話の流れに沿った自然な配置や広告の顕在性が評価された。しかし、助言と広告を同一のポットが提示することで応答の独立性が曖昧になり、特にカウンセリングのような文脈では信頼性低下や違和感を生むと報告された。

#### 6.2.3 AdBot

必要に応じて AdBot に切り替え、広告内容を深掘りできる点や、通常の MainBot との対話が侵入度の低い形で保たれる点が好意的に評価された。また、発話者ラベルや UI 上の区別により、助言と広告の境界がわかりやすく、役割や情報源の分離が知覚しやすい点が評価された。一方で、一部の参加者は AdBot が比較的目立たず見落とされやすいと指摘しており、呼び出しやすさや可視性を改善する余地が示唆された。

## 6.3 設計上の示唆

### 6.3.1 MainBot と AdBot の境界の明確化

本研究では、助言を行う MainBot と広告を提示する AdBot の情報源と責任帰属を恒常的に分離し、ラベル・アイコン・レイアウトで可視化した (図 2, 3)。実験では、AdBot 条件は発話内広告より高い信頼を示し、参加者のコメントからも「境界の曖昧さが信頼低下を招く」との指摘があり、境界の恒常的な可視化の妥当性が裏付けられた。

この分離はネイティブ広告への防御策としても機能することが予想される。一般に、非商業的コンテンツを模倣し情報源を不明瞭にする形式は欺瞞的であるとされ、商業性の明確かつ顕著な開示が求められている [28]。さらに、スポンサー付きや広告であることを示す明示的のマーカと適切かつ顕著な配置は、説得知識を高め [24]、情報源の混乱を減少させる [29] など、ブランドや媒体評価、信頼判断に体系的な影響を及ぼすことが知られている [12, 25]。本システムにおける恒常的な話者分離とラベリングは、広告性を明示し、誤帰属のリスクを低減できるといえるだろう。

### 6.3.2 顕在性と忌避感のトレードオフ

バナーブラインドネスを軽減し、ユーザの広告への関心を効果的に喚起する提示デザインは重要である [20, 21] 一方、顕在性の上昇は忌避感の増大を伴いがちである [30]。本研究でも、とくにカウンセリングでその傾向が確認され

表 4 総合評価にて収集したコメントの要約 (一部抜粋) .

システム	良い点/悪い点	各クラスタの要約文 (GPT-4.1 により生成)
バナー広告	良い点	<ul style="list-style-type: none"> <li>・ AI が押し付けている感じがせず、押し付けがましくなかった。</li> <li>・ 本文から明確に分離されており読みやすかった。</li> </ul>
バナー広告	悪い点	<ul style="list-style-type: none"> <li>・ 表示が控えめで、意識的に見ようとしないうりクリックしない。</li> <li>・ 応答テキストと特に関連性のない広告が時々表示された。</li> </ul>
発話内広告	良い点	<ul style="list-style-type: none"> <li>・ 会話の流れを妨げにくいと感じた。</li> <li>・ 会話のトピックに完璧に合致した広告が表示されることがあった。</li> </ul>
発話内広告	悪い点	<ul style="list-style-type: none"> <li>・ 非常に押し付けがましく感じた。</li> <li>・ すべての返信が広告のように感じられ、最後まで読む意欲が完全に失せた。</li> </ul>
AdBot	良い点	<ul style="list-style-type: none"> <li>・ 広告ポットとして独立してやり取りできるのが便利だった。</li> <li>・ 興味のある商品を深く調べられた。</li> </ul>
AdBot	悪い点	<ul style="list-style-type: none"> <li>・ アシスタントとポットの両方に質問できる点が煩雑であった。</li> <li>・ 広告を全く気にせず会話できてしまう。</li> </ul>

た。特に信頼性が重要な場面では、文脈に応じて顕在性と忌避性の境界で運用点を調整できる、細粒度かつ指標駆動の UI 最適化が求められる。

#### 6.4 制約と今後の課題

本研究では以下の制約が挙げられる。第一に、対象としたシナリオと広告ドメインが限定的な点である。本研究では製品推薦とカウンセリングの2文脈のみを対象に、約3,000件の広告アイテムのみを用いて評価を行った。この規模は現実の商用システムよりも小さく、システムへの信頼や広告受容度といった指標が多様な文脈でどのように変化するかについては十分に検証できていない。第二に、限られたターン数の単一セッション対話を前提としており、役割分離が長期的なユーザ体験や信頼形成に及ぼす影響は明らかではない。特に、提案手法においては他形式よりも対話者数が増えるため、それに伴うデメリットも考慮する必要がある。第三に、ユーザスタディがロールプレイベースのシミュレーションであった点である。実際の購買意図が伴わなかったため、エンゲージメントの機会が制限され、有用性や推薦精度の差が小さくなった可能性がある。

外的妥当性の強化に向け、今後の研究ではフィールド環境における逐次的な評価を進める。まず小規模なパイロット導入から開始し、タスクや広告ドメインを拡張し、マルチセッション利用を含めた長期評価を行う。また、調査票に基づく評価を補完するオンライン指標として、クリック率 (CTR)、再訪率、対話継続率などを測定し、AdBot 導入効果をより精緻に定量化する。これらを通じ、役割分離型アーキテクチャの有効範囲と限界を明確化し、実環境展開に向けた設計指針を示す。

## 7. おわりに

本研究では、広告を統合したチャットポットにおける信

頼性低下に対処するため、助言を担う *MainBot* と広告を担う *AdBot* を分離する二エージェント方式を提案した。予備実験では、対話ベースの方式がバナー広告よりもユーザの嗜好に適合した推薦が可能であることを示した。また、ユーザスタディにおいては、*AdBot* が発話内広告と比較してアシスタントの信頼性を向上させた一方で、広告の有用性については有意な差は見られなかった。これらの結果を総合すると、信頼が重視される場面において、本アプローチは有力な設計選択肢となり得ると考えられる。今後の研究では、段階的なフィールドスタディを通じて提案手法の効果を検証し、実運用に向けた設計指針を精緻化する。

#### 大規模言語モデルの利用について

本稿の執筆過程において、生成 AI を用いた。著者自身の文章の編集のみに使用され、最終的な内容が盗用、誤表現、捏造、改ざんを含まないことを確認した。

#### 参考文献

- [1] Kaiser, C., Kaiser, J., Schallner, R. and Schneider, S.: A New Era of Online Search? A Large-Scale Study of User Behavior and Personal Preferences during Practical Search Tasks with Generative AI versus Traditional Search Engines, in *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25, New York, NY, USA (2025), Association for Computing Machinery.
- [2] Yang, Y., Urgo, K., Arguello, J. and Capra, R.: Search+Chat: Integrating Search and GenAI to Support Users with Learning-oriented Search Tasks, CHIIR '25, p. 57–70, New York, NY, USA (2025), Association for Computing Machinery.
- [3] Liu, S., Hu, Y., Tian, Z., Jin, Z., Ruan, S. and Mao, J.: Investigating Users' Search Behavior and Outcome with ChatGPT in Learning-oriented Search Tasks, in *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, SIGIR-AP

- 2024, p. 103–113, New York, NY, USA (2024), Association for Computing Machinery.
- [4] Burke, M., Gorman, N., Nilsen, E. and Hornof, A.: Banner ads hinder visual search and are forgotten, in *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, p. 1139–1142, New York, NY, USA (2004), Association for Computing Machinery.
  - [5] Lee, J. and Ahn, J.-H.: Attention to Banner Ads and Their Effectiveness: An Eye-Tracking Approach, *Int. J. Electron. Commerce*, Vol. 17, No. 1, p. 119–137 (2012).
  - [6] Liu, Y. and Shrum, L. J.: What is Interactivity and is it Always Such a Good Thing? Implications of Definition, Person, and Situation for the Influence of Interactivity on Advertising Effectiveness, *Journal of Advertising*, Vol. 31, No. 4, pp. 53–64 (2002).
  - [7] Ghose, A. and Yang, S.: An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets, *Manage. Sci.*, Vol. 55, No. 10, p. 1605–1622 (2009).
  - [8] Zelch, I., Hagen, M. and Potthast, M.: A User Study on the Acceptance of Native Advertising in Generative IR, in *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*, CHIIR '24, p. 142–152, New York, NY, USA (2024), Association for Computing Machinery.
  - [9] Hajiaghayi, M., Lahaie, S., Rezaei, K. and Shin, S.: Ad Auctions for LLMs via Retrieval Augmented Generation, in *The Thirty-eighth Annual Conference on Neural Information Processing Systems* (2024).
  - [10] Schmidt, S., Zelch, I., Bevendorff, J., Stein, B., Hagen, M. and Potthast, M.: Detecting Generated Native Ads in Conversational Search, in *Companion Proceedings of the ACM Web Conference 2024*, WWW '24, p. 722–725, New York, NY, USA (2024), Association for Computing Machinery.
  - [11] Xiao, L., Ma, J., Dong, X. L., Martínez-Gómez, P., Zalmout, N., Zhang, C., Zhao, T., He, H. and Jin, Y.: End-to-End Conversational Search for Online Shopping with Utterance Transfer, in Moens, M.-F., Huang, X., Specia, L. and Yih, S. W.-t. eds., *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3477–3486, Online and Punta Cana, Dominican Republic (2021), Association for Computational Linguistics.
  - [12] Eisend, M., Reijmersdal, van E. A., Boerman, S. C. and Tarrahi, F.: A Meta-Analysis of the Effects of Disclosing Sponsored Content, *Journal of Advertising*, Vol. 49, No. 3, pp. 344–366 (2020).
  - [13] Liu, J.: ChatGPT: perspectives from human–computer interaction and psychology, *Frontiers in Artificial Intelligence*, Vol. Volume 7 - 2024, (2024).
  - [14] Liang, Y., Wu, Z., Zhang, F., Song, D. and Huang, H.: How Users Interact with Generative Information Retrieval Systems: A Study of User Behavior and Search Experience, in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '25, p. 634–644, New York, NY, USA (2025), Association for Computing Machinery.
  - [15] Bednarczyk, L., Reichenpfeiffer, D., Gaudet-Blavignac, C., Ette, A. K., Zaghir, J., Zheng, Y., Bensahla, A., Bjelogrić, M. and Lovis, C.: Scientific Evidence for Clinical Text Summarization Using Large Language Models: Scoping Review, *J Med Internet Res*, Vol. 27, p. e68998 (2025).
  - [16] Yuan, A., Garcia Colato, E., Pescosolido, B., Song, H. and Samtani, S.: Improving Workplace Well-being in Modern Organizations: A Review of Large Language Model-based Mental Health Chatbots, *ACM Trans. Manage. Inf. Syst.*, Vol. 16, No. 1 (2025).
  - [17] Noy, S. and Zhang, W.: Experimental evidence on the productivity effects of generative artificial intelligence, *Science*, Vol. 381, No. 6654, pp. 187–192 (2023).
  - [18] Jung, Y., Chen, C., Jang, E. and Sundar, S. S.: Do We Trust ChatGPT as much as Google Search and Wikipedia?, in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA (2024), Association for Computing Machinery.
  - [19] Sun, X., Ma, R., Zhao, X., Li, Z., Lindqvist, J., Ali, A. E. and Bosch, J. A.: Trusting the Search: Unraveling Human Trust in Health Information from Google and ChatGPT (2024).
  - [20] Benway, J. P. and Lane, D. M.: Banner Blindness: Web Searchers Often Miss "Obvious" Links (1998).
  - [21] Resnick, M. and Albert, W.: The Impact of Advertising Location and User Task on the Emergence of Banner Ad Blindness: An Eye-Tracking Study, *International Journal of Human–Computer Interaction*, Vol. 30, No. 3, pp. 206–219 (2014).
  - [22] Broeck, Van den E., Zarouali, B. and Poels, K.: Chatbot advertising effectiveness: When does the message get through?, *Comput. Hum. Behav.*, Vol. 98, No. C, p. 150–157 (2019).
  - [23] Christakopoulou, K., Radlinski, F. and Hofmann, K.: Towards Conversational Recommender Systems, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, p. 815–824, New York, NY, USA (2016), Association for Computing Machinery.
  - [24] Wojdyski, B. W.: The Deceptiveness of Sponsored News Articles: How Readers Recognize and Perceive Native Advertising, *American Behavioral Scientist*, Vol. 60, No. 12, pp. 1475–1491 (2016).
  - [25] Reijmersdal, E. A. V., Brussee, E., Evans, N. and Wojdyski, B. W.: Disclosure-Driven Recognition of Native Advertising: A Test of Two Competing Mechanisms, *Journal of Interactive Advertising*, Vol. 23, No. 2, pp. 85–97 (2023).
  - [26] Mayer, R. C., Davis, J. H. and Schoorman, F. D.: An Integrative Model of Organizational Trust, *The Academy of Management Review*, Vol. 20, No. 3, pp. 709–734 (1995).
  - [27] Hou, Y., Li, J., He, Z., Yan, A., Chen, X. and McAuley, J.: Bridging Language and Items for Retrieval and Recommendation, *arXiv preprint arXiv:2403.03952* (2024).
  - [28] Federal Trade Commission, : Enforcement Policy Statement on Deceptively Formatted Advertisements (2015), Accessed 2025-09-08.
  - [29] Boerman, S. C., Reijmersdal, van E. A. and Neijens, P. C.: Sponsorship Disclosure: Effects of Duration on Persuasion Knowledge and Brand Responses, *Journal of Communication*, Vol. 62, No. 6, pp. 1047–1064 (2013).
  - [30] Edwards, S. M., Li, H. and Lee, J.-H.: Forced Exposure and Psychological Reactance: Antecedents and Consequences of the Perceived Intrusiveness of Pop-Up Ads, *Journal of Advertising*, Vol. 31, No. 3, pp. 83–95 (2002).