

【田中頼人特別研究】

第12回レポート（中間発表）

2301330039：安田直也



AI検索サービスにおける企業優遇バイアス：市場競争への潜在的リスク

※バイアス：LLMの根拠のない、あるいは偏った出力

なぜこのような研究を行うのか？

皆さんは普段、Googleなどの検索エンジンで情報を調べますよね。

でも、もしその検索結果が、特定の企業だけを優遇するように操作されていたとしたら、どう思いますか？

特定の企業ばかりが検索結果の上位に表示されたり、目立って表示されるようになれば、他の企業は不利になり、**私たち消費者が選べる商品やサービスが限られてしまう**かもしれません。

この研究では、**ChatGPTやPerplexityといったAIを使った新しい検索サービスが、このような『企業優遇バイアス』を持つ可能性とその影響**について調べていきます。

なぜこのような研究が必要か？

現在、AI技術は急速に進化しており、規制や倫理的なガイドラインがそのスピードに追いついていないのが現状です。

AI検索サービスがどのようにバイアスを生み出し、それが市場競争や社会全体にどのような影響を与えるかを明らかにすることは、今後のルール作りや技術の正しい進化の方向を決めるために欠かせません。

また、AIの透明性と公平性を確保するための対策を講じなければ、技術がもたらす恩恵が一部の企業だけに集中し、社会全体の利益を損なう結果になりかねません。この研究は、私たちが直面するこの課題を解決するための重要な一歩です。

このような研究が求められる社会的・理論的な背景は何か？

AI検索サービスは便利ですが、どのように結果が決まっているのかが分かりにくいという課題があります。この「ブラックボックス化」によって、特定の企業が優遇されても、多くの人が気づきにくい状況があります。

特定の企業が優遇されると市場競争が不公平になり、新規参入企業や中小企業が市場に参入しづらくなるリスクがあります。また、消費者にとっても、多様な選択肢が失われ、最良の判断を行うことが難しくなるでしょう。

こうした偏りが続くと、社会全体の公正性が損なわれ、企業間の競争バランスが崩れるリスクもあります。この研究を通じて、AI検索サービスが与える競争市場への潜在的リスクについて考察し、政策提言や倫理的ガイドラインの形成に寄与することを目指します。

これまでにどのような先行研究があるか？

- ・ LLMで企業名や企業属性が金融テキストの感情評価にどのように影響するかを調査したものがあり企業固有のバイアスを評価したものとしては唯一の研究
- ・ LLMが特定の経済圏や出身国に基づくブランドバイアスを持つことを示した研究も存在し、商品ブランドに関連するバイアスに焦点を当てている
- ・ LLMを使用した推薦システムで音楽、映画、書籍の推薦における人口統計的・文化的バイアスがあることを示しており、AI検索サービスも情報の推薦を行うため先行研究として参考になる
- ・ 最近の研究では、Google AI Overviewsが白人の遺伝的優位性を示唆するようなIQに関する偏った回答を生成したとの報告もあり、copilot AIとPerplexityも同様にそういった偏った統計情報を参照しているとの指摘がある

先行研究からどのような知見を得られたか？

- LLMにはポジティブやネガティブといった感情評価を持っていること
- LLM側にロールを与えたり、対話する人間側の情報（文化的、人種的、地理的な情報）を設定することでバイアスが発現しやすくなること
- バイアスの定量化、評価指標
- AIの公平性と企業の責任についての論文が近年は多く、注目度が高いこと

得られてきた知見に対して、まだ不足していることは何か？

- 企業優遇バイアスの研究自体がほぼない
- LLMとバイアスに関する研究が多く、AI検索サービスとバイアスを結びつけている研究はあまりない
- 日本語での対話で調査した研究、論文がまだ少ない
- 企業優遇バイアスに関するデータセットがない
- 競争市場への影響についての考察

不足している部分を埋める意義は何か？

- 企業優遇バイアスや、AI検索サービスのバイアスは未開拓領域の研究であるため新たな知見を提供できる
- 日本語での対話に基づく研究が少ないため、日本語環境における企業優遇バイアスの特性を明らかにすることができる
- 企業優遇バイアスに関するデータセットを新たに構築することで、今後の研究や実証分析の基盤を提供できる
- AI検索サービスの企業優遇バイアスが競争市場に与える影響を考察することで、公正な市場競争を守るための政策提言や倫理指針の作成に貢献できる

どんな問題を解決するために、自分は何をやろうとしているのか？ 研究を通じて何を明らかにできるか？

AI検索サービスが特定の企業や商品、サービスを偏って推奨することで、人々の購買行動にバイアスをかけ、公正な競争市場を損なう問題がある。この問題を解決するため、AI検索サービスにおける企業優遇バイアスを明らかにし、それが競争市場に与える影響を考察する。

また、従来の手動による検証に加え、企業優遇バイアスの調査を一部システム化することで、効率的かつ客観的な分析を目指す。

どのような手法で研究を行うのか？

①感情評価によるバイアス検出

商品・サービスに関する簡単な例文を作成し、商品・サービス名をマスクしたプロンプトとしていないプロンプトでの感情評価を1～5で出してもらう。

マスク有りの感情評価とマスク無しの感情評価の差をその商品・サービスに関する企業優遇バイアスの指標とする。
同様に企業のトピック（環境保護活動など）に関しての例文で企業名のマスク有無のプロンプトで感情評価を比較し、
こちらも企業優遇バイアスの指標とする。

②オススメ順ソートによるバイアス検出

あるサービスカテゴリ（例えばクラウドサービスやストリーミングサービス）に関して、
オススメ順でサービス名を列挙してもらう（AWS、Azure、Google Cloud…）プロンプトを実行する。
他のAI検索サービスとの比較や売上順との比較をする。

③Google検索との比較によるバイアス検出

企業名や企業のトピックに関してGoogle検索を行いAI検索サービスの検索結果と比較する。
表示されるリンクの上位にある公式情報の割合の差、ネガティブ情報の差を比較する。

※これらはPerplexityなどのAPIがあるAI検索サービスの場合はAPIを使用してコード実行により検証する
APIがないサービスは手動でプロンプトを実行する

なぜそのような手法を選んだのか？

①感情評価によるバイアス検出

先行研究により手法が確立されており、準備や実行が比較的容易であるため、再現性と効率性の観点から適している。

②オススメ順ソートによるバイアス検出

実験的にプロンプトを試した結果、多くのAI検索サービスで順位付けが可能であることを確認したため、バイアスの検出に有効であると判断した。

③Google検索との比較によるバイアス検出

既存の検索サービスとの比較を通じて、AI検索サービスがどの程度偏りを持っているかを明確にすることが重要であると考えたため。

どのようなデータを手に入れたか？

プロンプトテンプレートデータセット：

感情評価用、オススメ順取得用、検索結果比較用の標準化されたプロンプト

感情評価に基づく企業バイアスデータセット：

企業・商品・サービスごとのバイアス指標

推奨順位データセット：

各AI検索サービスでの商品・サービスのオススメ順位リスト

今の時点で、どのようなソフトウェアを試作できたか？

「感情評価によるバイアス検出手法」に基づき、サービス名のマスク有無それぞれに対応したプロンプトを作成し、PerplexityのAPIを使用して商品・サービスのバイアスを検出する仕組みをGoogle Colabで試作した。

次のステップとして結果のスプレッドシートへの出力と、ChatGPTのAPIへの横展開を計画している。

どのような方針で評価しようとしているか？

①先述の手法ごとに企業優遇バイアスの指標を算出し、AI検索サービス間での差異を比較するために、ヒートマップ形式で可視化を行う。

②オススメ順や企業のトピックに関する検索結果を提示する前後で五段階評価のアンケートを実施し、消費者の印象や購買意欲に与える影響を評価する。

③AI検索サービスとGoogle検索での結果を比較し、表示されるリンクの上位にある公式情報の割合の差、ネガティブな記事のリンク数の差などで評価する。

④上記の結果を基に、特定企業やサービスカテゴリーがバイアスを受ける割合を算出する。

バイアス割合① = 特定企業のスコア差 / 全体のスコア差合計

バイアス割合② = 特定企業の上位順位回数 / 対象の全企業の上位順位回数合計

バイアス割合③：

$$\text{バイアス割合} = \frac{\text{公式情報の割合}_{AI} - \text{公式情報の割合}_{Google}}{\text{公式情報の割合}_{Google}}$$

どのような方針で評価しようとしているか？（続き）

⑤バイアス割合を市場シェアに反映

算出したバイアス割合（例: 特定企業のスコア差や順位優遇の割合）を基に、**仮想的な市場シェアの偏りを計算する。**
具体的には、各企業の市場シェアを元データとし、バイアス割合を加味して調整しその市場シェアを基に市場集中度（HHI）を算出。

$$HHI = \sum_{i=1}^N (\text{調整後市場シェア}_i)^2$$

バイアスを加味したHHIと元の市場シェアを基に算出したHHIを比較し、市場競争への影響を評価。
HHIの変化が大きいほど、AI検索サービスが市場集中を強める可能性を示唆。

バイアス割合によりシェアが減少する企業やカテゴリーを特定し、新規参入企業が不利になる市場状況を考察する。

先行研究となる論文・書籍

中川 慧、平野 正徳、藤本 悠吾：大規模言語モデルを活用した金融センチメント分析における企業固有バイアスの評価、第21回テキストアナリティクス・シンポジウム、vol.124, no.173, NLC2024-15、pp.81-86（2024年9月3日）

Kamruzzaman, M.; Nguyen, H. M.; Kim, G. L.: "Global is Good, Local is Bad?": Understanding Brand Bias in LLMs, arXiv preprint, arXiv:2406.13997, (2024)

Sakib, S. K.; Das, A. B.: "Challenging Fairness: A Comprehensive Exploration of Bias in LLM-Based Recommendations," arXiv preprint, arXiv:2409.10825, (2024)

WIRED: "Google, Microsoft, and Perplexity Are Promoting Scientific Racism in Search Results," WIRED, <https://www.wired.com/story/google-microsoft-perplexity-scientific-racism-search-results-ai/>,（アクセス日: 2025年1月3日）

手法や評価方法の参考にしたもの

FLTech Blog: "LLM Bias Diagnosis: 大規模言語モデルのバイアス診断と緩和に関する最新動向," FLTech Blog, <https://blog.fltech.dev/entry/2024/03/22/LLM-Bias-Diagnosis-ja>,（アクセス日：2025年1月3日）