Monetization Risks in AI Systems: A Bing Copilot case study

Submission to the European Commission

Ilan Strauss, Tim O'Reilly, Mariana Mazzucato, and Julius Schulte *

April 2024

Abstract

We conduct a descriptive analysis and simple experiments on Microsoft's AI Copilot LLM – a service that provides answers in a comparable manner to a Search results engine. We interrogate the nature of Copilot's generated results, characterized by notable use of advertising to monetize user attention already, and compare it to Microsoft's traditional Bing search engine results on a series of e-commerce product queries. We find that the underlying sources from which the generated results in Copilot ostensibly come from (the footnote links) differs substantially from the websites listed in Bing's traditional search results, but for no apparent reason. A lack of algorithmic transparency and explainability in Bing's Copilot opens up room for providing suboptimal text results in return for paid placement (advertising). We highlight that use of retrieval-augmented generation (RAG) by LLMs, to draw on a wider body of relevant material, creates significant monetization opportunities for LLMs. To avoid past mistakes of digital platforms, advertising output needs to be clearly demarcated, limited in scope, and regulated for quality. We also recommend that LLM generated results should be clearly demarcated from other types of algorithmic results. Such interventions are important to undertake early, since the room for insidious monetization through LLMs will significantly increase alongside growing user trust and reliance on them.

*Thank you to Julius Schulte, data scientist at World Economic Forum, for leading the empirical work (Bing case study) in this submission. This was submitted as part of a European Commission's call for contributions on "Competition in AI and Virtual Worlds": https://competition-policy.ec.europa.eu/about/europes-digital-future_en. Ilan Strauss is the Director of AI Disclosures Project at the Social Science Research Council and Senior Fellow at UCL's Institute for Innovation and Public Purpose (IIPP). Tim O'Reilly is the Founder, CEO, and Chairman of O'Reilly Media, Co-Director of the AI Disclosures Project, and Visiting Professor of Practice at UCL IIPP; Mariana Mazzucato is Founding Director and Professor in the Economics of Innovation and Public Value at UCL IIPP; Julius Schulte provided research assistance for this project. Disclaimer: The views expressed in this paper are solely those of the authors and do not necessarily reflect the views or opinions of their employer. Corresponding author: Ilan Strauss (i.strauss@ucl.ac.uk), 11 Montague St, London WC1B 5BP, United Kingdom.

1 Introduction

Following the European Commission's call for contributions, we consider, within a competition framework (European Commission 2024): "How will generative AI systems and/or components, including AI models likely be monetized, and which components will likely capture most of this monetization?".

We conduct a series of basic experiments on Microsoft's AI Copilot large language model (LLM) to explore possible avenues for it's monetization, including to what extent it is already being monetized. Microsoft's AI Copilot LLM (previously called Chat) is powered by OpenAI's GPT 4.0 model and can provide the user with generated answers in a comparable manner to Microsoft's traditional Search results engine. Both services are accessed via the same web page - though Search remains the default,[1] and although some integration of the two products has taken place. Our initial findings are as follows:

Descriptively, Copilot already monetizes user attention by promoting paid links above organic links. Paid links appear when the user looks to click for more information, or in this case to make a purchase. Organic footnoted purchase links are placed below the paid links on the screen, which appear above to the right as a pop-up. Copilot integrates what appears to be "traditional" vertical shopping search ads in-between its LLM generated content. It is not clear in Copilot's generated answer (it's "result") if the organic visual information provided right at the bottom of its long answer comes from Bing Search or an LLM. Lastly, in Copilot, arguably the most useful organic information, enabling more effective product comparisons, appears at the bottom of the screen for e-commerce shopping.

Next, we run a set of queries looking to buy e-commerce products on both Copilot and Search. We find that the actual underlying sources used ostensibly to generate results in Copilot (footnote links) differs substantially from the website links displayed in Bing's traditional search results. Bing Copilot's responses cite Bing search as a data source only on occasion, suggesting a possibility for companies' websites to optimize their presence without disclosure of the mechanism to the user. The reasons for these differences are unclear and highlight a lack of algorithmic transparency and explainability in Bing's Copilot. This opens up room for providing suboptimal text results in return for paid placement (advertising and sponsored results).

Although such monetization risks in Copilot may appear similar to how Google, Amazon, and others have degraded their search results with ads in the past, this is arguably of a different potential magnitude because these algorithms are more powerful cognitively and so trust in them may become far

[1] See: www.bing.com.

greater. Moreover, these algorithms are black boxes, such that explaining to the user how they work - or even guaranteeing to the user that they are in fact producing the best results - may not even be possible.

To avoid previous regulatory missteps, which permitted digital platforms to monetize their algorithmic outputs excessively and opaquely (O'Reilly, Strauss, and Mariana Mazzucato 2024; Strauss et al. 2024; Mariana Mazzucato, Strauss, et al. 2023), we make the following broad recommendations:

1. Use of retrieval-augmented generation (RAG) by LLM systems (P. Lewis et al. 2020) pose significant risks for opaque monetization. Any inclusion of paid information into a RAG needs to be disclosed by the company in some form.

2. More broadly, including paid information in an AI generated result, its source link, or in its training data poses risks to developing an equitable ecosystem of third-party content creators and AI entities, which helps reward the creators of the most relevant information for the user.

3. Advertising output needs to be clearly demarcated, limited in scope, and regulated for quality. Low quality advertising has created significant harms on the internet today, which risk being replicated in the AI system upgrade versions of these products, as well through entirely new products.

4. LLM generated search and other text results should be clearly demarcated as such in Search and other services given that the explainability of their results is low, making external verification that they are showing the results which are most beneficial to the user difficult or impossible.

5. Algorithmic transparency and disclosures made by AI companies in their 10-K and quarterly reports is an important basis for fostering effective competitive regulations. This supports policing abusive behaviour *ex-ante*, and helps regulators develop a better understanding of a market's structure and its core institutions (O'Reilly 2023; Strauss et al. 2024).

The above results are tentative, based on small datasets and a preliminary exploration of some of the data – not all of which we have had the time to write up here. Moreover, Microsoft's product offerings, based on OpenAI's GPT 4.0, are constantly changing – not just in name, but in shape, and in how and to what extent they are monetized.

Still, the above interventions are important to undertake early, since the room for opaque and harmful monetization through LLMs is will become great user trust and reliance on these AI systems grow (O'Reilly, Strauss, and Mariana Mazzucato 2024). Ingrained algorithmic habits by users, combined with "algorithmic authority", underpinned the ability of Big Tech to monetize their algorithms previously.

Looking forward, it only takes a very small leap of the imagination to envisage a future whereby LLMs generate advertising just as they generate organic content, that the two will become increasingly intertwined, that telling the difference between the two for the consumer may become exceedingly difficult, and that this may impose great time and cognitive costs on the user.

-*-

Our previous work at University College London's *Institute for Innovation and Public Purpose* outlined the centrality of algorithms in managing multi-sided marketplaces and in extracting above-normal profits from third-party firms, through allocating user attention excessively to sub-optimal information (O'Reilly, Strauss, and Mariana Mazzucato 2024). The fact that such rents can persist speaks not just to barriers to entry and sticky user behaviour, but also to a user's inability to constantly inspect and compare the information shown to them by algorithms (Rock et al. 2023). Contrary to Neoclassical theories, time and cognition are costs to users, users satisfice and do not know the costs and benefits of additional search in advance (Strauss et al. 2024). This allows for Big Tech's algorithms to exploit such heuristic behaviour (O'Reilly, Strauss, and Mariana Mazzucato 2024). In Rock et al. (2023), we note that practically, digital market power is monetized when:

> *A dominant platform can use advertising as a price substitute (E. Hovenkamp 2018) to get users to pay with more of their attention (i.e., time) than what would prevail in a competitive market, without information and behavioural imperfections.*[2] *We call this "attention rent" (O'Reilly, Strauss, and Mariana Mazzucato 2024; Mariana Mazzucato, Ryan-Collins, et al. 2023). For users, excessive advertising may lead to inferior product matching (Roth 2015), higher search costs (Areeda and H. Hovenkamp 2023; Ursu 2018),*[3] *and less variety due to considerable duplication of products in search results.*[...]
>
> *Advertising output above a certain level may also unfairly exploit the platform's third-party ecosystem of advertisers and/or suppliers. This follows from each side of a platform competing for the user's attention and the platform's algorithmic attention allocations deciding how value is allocated between these competing sides* (O'Reilly, Strauss, and Mariana Mazzucato 2024; Strauss et al. 2024). *When advertising crowds out organic results, it can create a dynamic in which suppliers are strongly incentivized to pay to receive user attention, rather*

[2] More advertising amounts to a higher effective shadow price paid by users (Baye and Prince 2020).

[3] Areeda and H. Hovenkamp (2023, Section 2023, Agreements Pertaining to Advertising and Related Dissemination of Product Information): "increased consumer search costs can yield higher pricing".

> *than to earn it competitively.*[...]
>
> *A platform's successful exploitation of its users through advertising relies on fairly persistent user click behaviour on advertising results. Otherwise more advertising output will not be profitable and instead will lead to a decline in total profits as advertising prices and/or product sales fall.*

Our broad argument is that such algorithm exploitation via paid information can and will evolve for the use cases and business model of generative AI systems. The exact nature may differ if scale effects continue to be more important than network effects. But advertising and insidious monetization is likely to degrade the quality of information in AI systems, platforms, and services, in a comparable manner to how Big Tech's search and recommendation algorithms, in search of monopoly profits, degraded algorithmic quality and disregarded user preferences on Amazon (Rock et al. 2023), Google (O'Reilly, Strauss, and Mariana Mazzucato 2024), and Facebook (Mariana Mazzucato and Strauss 2024).[4] It is important to try to shift the economic incentives underpinning today's AI innovations, away from risky uses of AI technology and towards open, accountable, AI algorithms that support and disperse value equitably, and are long-term orientated in nature (Mariana Mazzucato and Strauss 2024; O'Reilly 2024).

2 Case Study: Bing Copilot and attention monetization

Drawing on Herbert Simon (O'Reilly, Strauss, and Mariana Mazzucato 2024), McCurley (2007), an ex-Google software engineer, notes that the currency of online markets is attention:

> *The scarce resource in this [world wide web] market is not the information itself, nor is it money to buy the information. The scarce resource that plays the role of currency in this market is attention.* [...] *Authors compete for attention because attention has value. The essence of advertising is to steer the attention of consumers toward products offered by a seller. When advertising succeeds, it is because it completes a transaction that turns attention into monetary value from sale of goods or services. Attention can often be converted into other forms of value, such as reputation.*[...] *A prerequisite for monetization of web resources is to garner attention.*

[4] For legal discussion of this, and with respect to key court cases, see: Strauss et al. (2024).

In order to understand the institutional mechanisms and processes that shape information markets online, therefore, we must understand the process by which user attention is matched and prioritized with a given piece of information by an LLM. This requires us to focus on information retrieval in LLMs and how this information is presented to (i.e. consumed by) the user. We explore these factors below using a cursory case study of how information is retrieved and presented by Bing's AI Copilot (formerly Bing Chat), which uses OpenAI's GPT 4.0 LLM, and we contrast this with the URL website results provided by Bing's classic Search engine service, based on more traditional Search technology.[5]

Our exact method is explained in the Appendix A in more detail and throughout the text. It involves first some descriptive visuals from Bing's Copilot generated results, followed by comparing the generated results shown by Copilot vs. Bing Search results, focusing first on the different order of website links in each – the underlying sources of the information drawn on by the two (footnote sources in Copilot and organic results in Search); and finally assessing the different ordering of these URLs in the two more systematically, computing the Jaccard similarity score of the two (Figure 5).[6]

Our preliminary **findings** are that:

- Bing's AI Copilot is already monetizing its LLM results by steering users seeking additional information, or looking to make a purchase, towards paid links. In some generated results, although Bing Copilot relies heavily on Bing Search data for its RAG, it exclusively suggests commercial URLs without presenting any Bing Search URLs.

- Bing Search and Bing's AI Copilot employ distinct methodologies to prioritize the relevance of information, as evidenced by the sequence in which URLs (footnote links in Copilot vs. organic website links in Search) are suggested as potential sources of information. This difference in prioritization influences how users allocate their attention and, consequently, impacts the value received by firms competing for user attention through Bing's platform.

- The differing sequence of information sources displayed in Bing Search and Bing's AI Copilot reveals a concealed and significant potential for monetization within user-facing generative AI systems.

- It is unclear what factors determine the sequence of results shown and sources used in each model service. This underscores the necessity for greater AI transparency and explainability in

[5] Of crawling, indexing, query processing, and a ranking algorithm - but which may also increasingly make use AI for query matching and results summarization.

[6] We do this only when the Copilot footnote links include "www.bing.com" as one of the footnote sources, which we can then search directly and scrape the links from.

the implementation of RAG and AI systems.

2.1 The retrieval-augmented generation (RAG) in LLMs

When trying to understand the opaque and potentially dangerous monetization of LLMs' generated information, how its information is chosen and then presented becomes of chief concern.

By way of background on how information is selected by LLMs, it is important to understand that, today, user interaction with a foundation model relies on information not simply from a trained and fine-tuned foundation model (FM), but also on additional context (information) brought into the model through retrieval-augmented generation (RAG). Current examples of RAG based chatbots include Bing Chat, Google's Bard, and OpenAI's GPT 4.0 (OpenAI 2023). When a user interacts with GPT 4.0 directly on OpenAI's website and the LLM says "searching Bing" this is part of the RAG phase, for example.

RAG is a data retrieval mechanism that selects relevant data to help answer a user query (e.g. "What is the best exercise to fix back pain?") based on how similar the information is to the user's query. RAG involves both an information retrieval and an augmentation phase used to enhance the model's base knowledge.[7]

RAG is valuable for queries where the model needs knowledge which is not included in the information the model has learned from (ibid.). But in theory, RAG can be subverted and used to include information opaquely, even when not strictly needed. For example, a RAG could be used to include additional information in a user's query based on an advertising-like bidding system, even when the LLM has the answer, or when the LLM's answer would be more accurate if it instead used an"organic" RAG mechanism. In the instance of advertising, the information shown to the user may be a function, at least partly, of who paid the most to be included in the RAG.

In theory, RAG should make the information more accurate and more relevant. But if subverted for financial gain, RAG has the potential to make information less accurate and less relevant. Moreover, this can have significant implications for fostering competitive market online, which should reward the best information with user attention and punish less relevant, inferior, information. As with the present

[7] How RAG Works (simplified): **Retrieval Phase**: When presented with a query or question, the RAG system first searches a large database or corpus of documents to find relevant information. This retrieval is typically performed using a vector-based search, where the query and documents are converted into vectors in a high-dimensional space, and similarity metrics are used to identify the best matches; **Augmentation Phase**: The retrieved documents or pieces of information are then fed into the LLM along with the original query. This process enriches the context available to the model, providing it with specific details or data points that were not part of its initial training set; **Generation Phase**: Armed with both the original query and the retrieved information, the LLM generates a response. This response is informed by the broader knowledge the model gained during its pre-training, as well as the specific, relevant information contained in the retrieved documents.

growing dominance of paid information in algorithmic output online, letting RAG be dominated by paid information takes advantage of user trust in AI and subverts a fair market mechanism which rewards competitive information (products, websites, content creators) with user attention – substituting it with one which rewards the highest bidder.

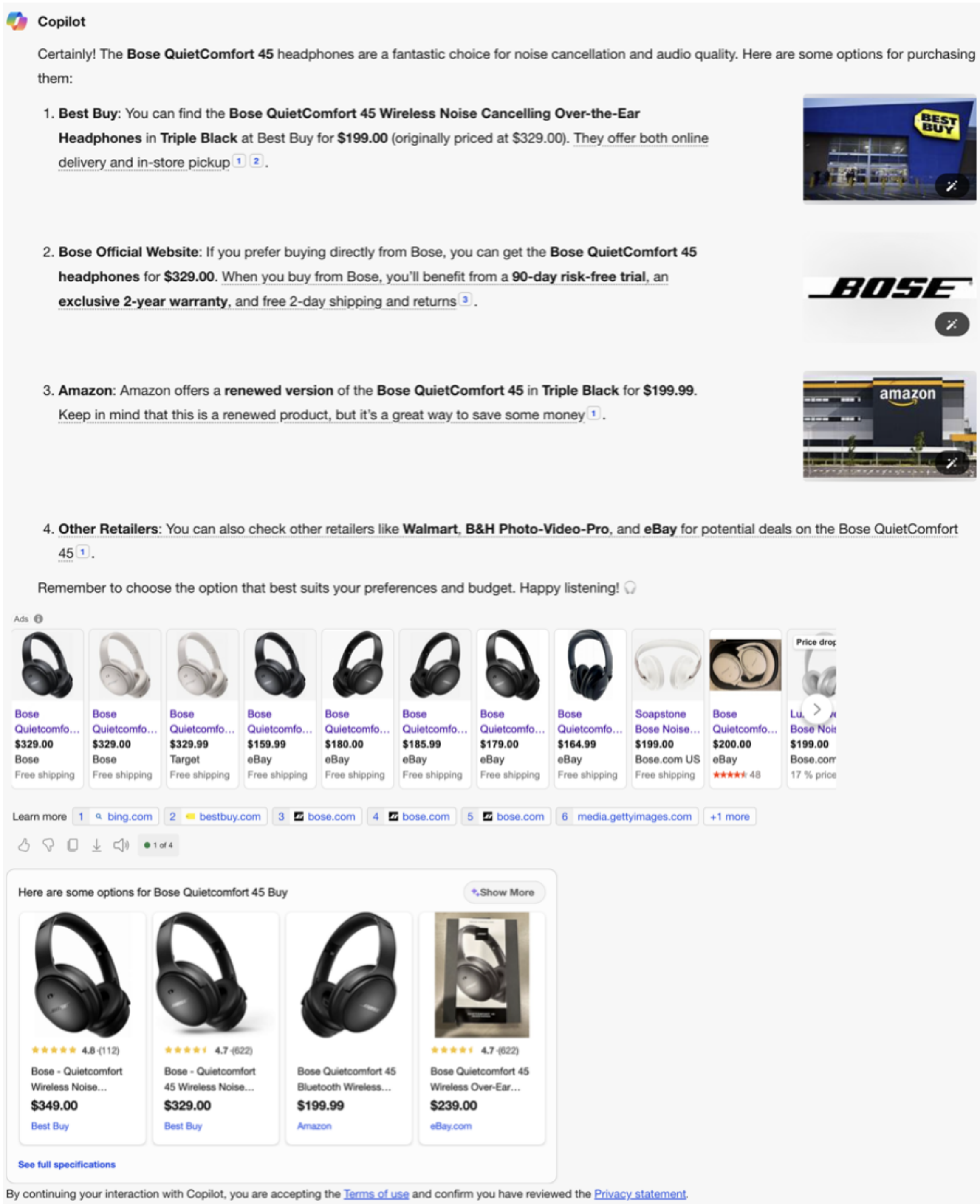
Monetizing LLMs through ads-based monetization may be an advantageous business strategy for larger AI companies with deeper pockets looking to gain market share over paid rivals. This means that it may help reinforce dominance or size effects. As network effects become more important to LLMs, perhaps with greater use of user data to enhance personalization, the ubiquity of this strategy may grow.

2.2 Bing Study

To descriptively explore Bing's Copilot we ask it the e-commerce query: "Where should I buy a Bose Quietcomfort 45?" headphone[8]. In turn, the user receives the response shown in Figure 1. To achieve this response, the user visits "www.bing.com" and clicks "Copilot" on the top left corner and one types inside of Copilot the above query. We use the normal ("balanced") AI model. The total response to this answer (as of the time of this research) is shown in Figure 1 below. The top grey Copilot answer is an LLM generated response with a numbered list. Below it are sponsored (possibly more traditional vertical search ads) followed by footnotes which connect to the initial LLM generated text above the sponsored ads. The third component at the bottom contains pictures which compare the prices and reviews of the given product sellers. It appears to be organic, and may or may not be LLM generated (in conjunction with a RAG). We break down each of these components in turn.

[8] Assuming a U.S. IP address.

Figure 1. Overview of Bing's Copilot Response to a Shopping Query.

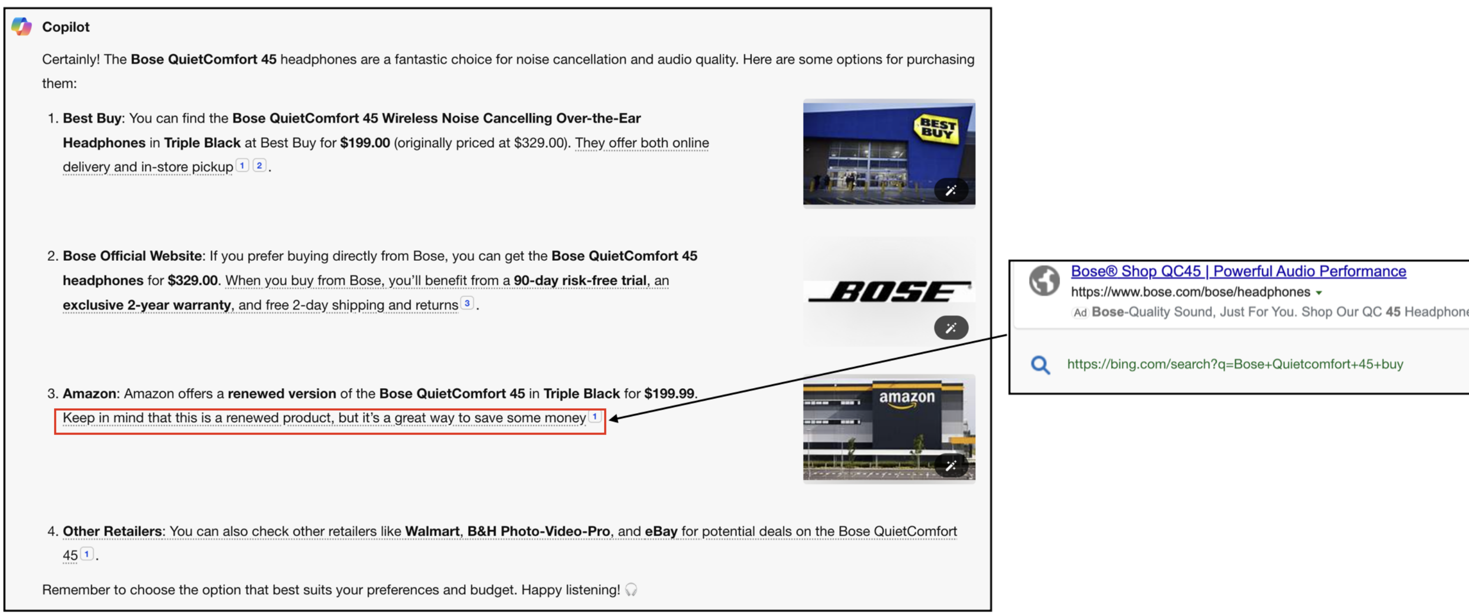


Note: *A typical Bing AI Copilot response to the query: "Where should I buy a Bose Quietcomfort 45?" headphone set, shows that advertising is already ubiquitous, directing user attention to paid links rather than organic links (08.03.2024).*

A typical Bing Copilot response in relation to our query can be divided into three parts: Part I (see Figure 2) of Bing's AI Copilot response involves an AI generated text summary, augmented by a RAG to find the actual information, and containing a list of places where the user can buy Bose Quietcomfort

45 headphones. These text suggestions are underlined with a dotted line that indicates one can hover over the text for more information. Hovering over the text displays a pop-up box (at the position of your mouse cursor) with an advertisement URL at the top of the box and a Bing search suggestion at the bottom of the box. Sometimes the Bing search suggestion box on the right is completely filled with paid advertising URLs, top and bottom. Note that in this example, the advertising result link is displayed more prominently above the organic Bing Search result link, while the word "ad" is barely visible, in keeping with present practice on platforms to barely distinguish an organic result from an advertising result.

Figure 2. Part I of Bing's AI Copilot Response to a Shopping Query.

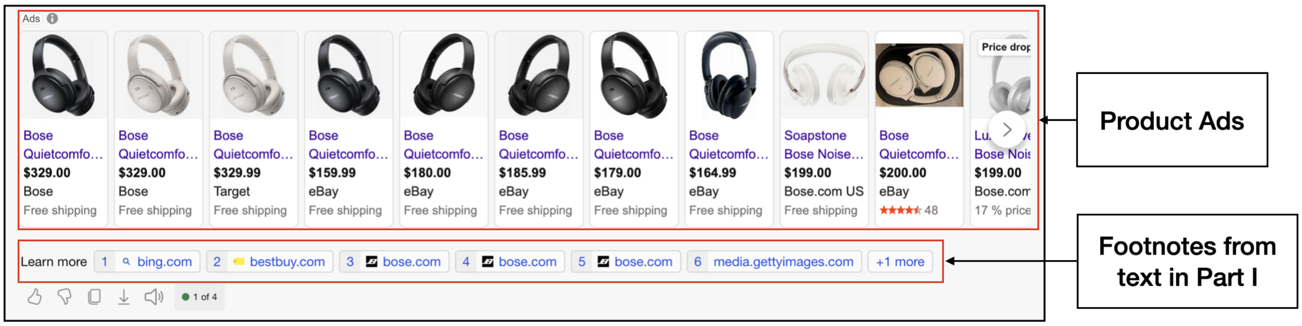


Note: *Part I of a typical Bing AI Copilot response to the query: "Where should I buy a Bose Quietcomfort 45?" headphone set, shows that advertising is already ubiquitous, directing user attention to paid links rather than organic links (08.03.2024).*

Figure 3 comes directly under Figure 2 on the screen and contains advertising and footnote links which relate to the LLM generated numbered list. The user does not have to click on anything else. The product ads in Figure 3 appear to be Bing Search ads, bid for in Bing's shopping vertical. It is unclear how the Bing Search ads relate to the co-pilot ads. These search ads appear below the LLM results but above the LLM footnotes, thereby helping intersperse them within the LLM results section - even though on a very rudimentary basis here.

At the end of each LLM summarized text response, the text in Figure 2 contains a footnote in Figure 3 that links to several information sources. The information sources contained in these footnotes are the same as the non-ad links suggested in Figure 2 when hovering over the text. The footnote links also contain the sources of the images that you can observe in Figure 2. The images have the potential to reinforce a certain message or brand.

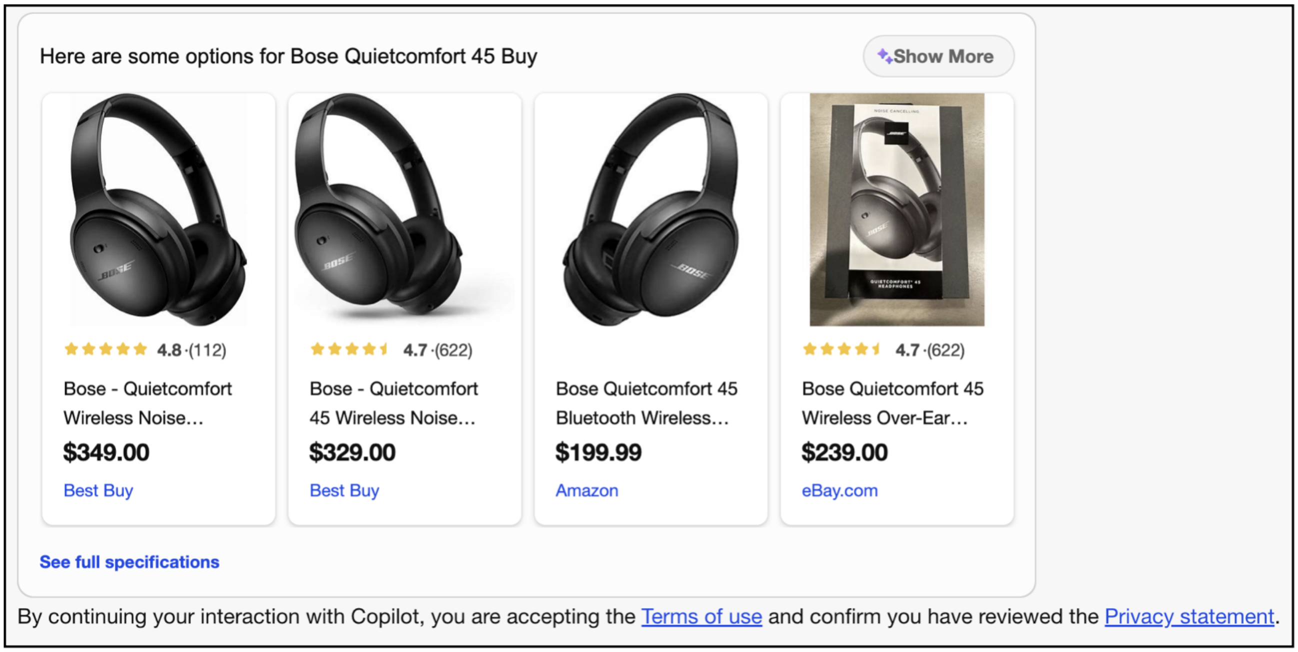
Figure 3. Part II of Bing's Copilot Response.



Note: *Part II of a typical Bing Copilot response to "Where should I buy a Bose Quietcomfort 45?" (08.03.2024).*

Lastly, Part III of the Bing Copilot response provides another product comparison that is not indicated as an ad (Figure 3). It is not labeled as a Copilot AI generated response or a Search result. It usefully facilitates user comparison of product features, such as price and user reviews, on an organic basis, yet it appears at the bottom of the screen. Clicking on "See full specifications" provides a scrollable table with the features of the product, which is likely an AI generated response taken from the websites. (In the scenario at hand it is always the same product, however, in the context of a more varied question this allows to compare different products.) Note that the *Bose* official site does not appear here even though it was the second recommended link in the Copilot response. Instead, eBay appears.

Figure 4. Part III of Bing's Copilot Response.



Note: *Part III of a typical Bing Copilot response to "Where should I buy a Bose Quietcomfort 45?" (08.03.2024).*

Next, Table 1 shows the order of the website links used in the results for Bing's Copilot vs. Bing's Search. We repeat the same query five times and show the results for: "Where should I buy a Bose Quietcomfort 45?". In the Bing Copilot ranking (formerly Bing Chat): Best Buy, Bose, and Bing consistently take the first three footnoted spots below the generated results and content. In contrast, looking at Bing's organic results (and using the query that is suggested by Bing Copilot's footnote to Bing Search): Best Buy, Amazon and Bose, followed by Target, are the first few organic websites listed. What can we draw from this? Firstly, aggregators play an indirect role in the rankings of Copilot through Bing Search. Copilot appears to provide more weight to Bose's direct site rather than to other aggregator shopping sites. Amazon plays a more direct role in Bing's Search rankings. And Best Buy appears at the top of both lists implying considerable commonality.

Table 1. Comparison of links in Copilot AI vs. Bing Organic Search for the same Query (Repeated five times)

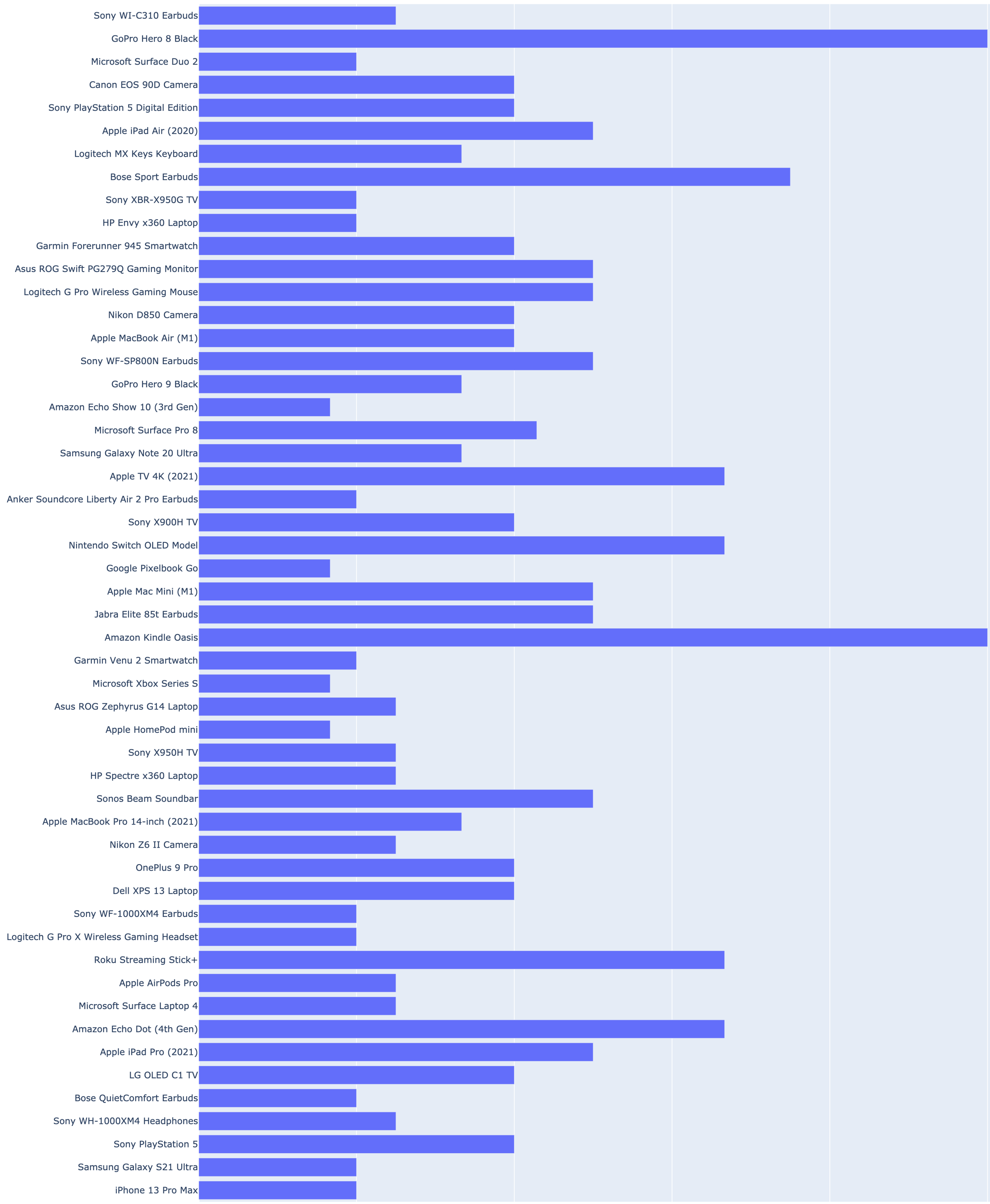
Copilot AI (footnote links order)	**Bing Search websites** (Organic order)
1. Best Buy	1. Best Buy
2. Bose	2. Amazon
3. Bose	3. Target
4. Bing	4. Bose
1. Best Buy	1. Best Buy
2. Bing	2. Amazon
3. Bose	3. Bose
4. Bose	4. Target
1. Bing	1. Best Buy
2. Best Buy	2. Amazon
3. Bose	3. Bose
4. Bose	4. Target
1. Bing	1. Bose
2. Best Buy	2. Best Buy
3. Bose	3. Amazon
4. Bose	4. Bose
1. Bing	1. Best Buy
2. Bing	2. Amazon
3. Best Buy	3. Bose
4. Best Buy	4. Target
5. Bing	5. Best Buy

Note: *The table compares the order of the URLs suggested by Bing Chat (in its footnotes) and Bing Search in its ordinary organic links listed for the query "Where should I buy a Bose Quietcomfort 45?". This is repeated several times. Showing only the minimum number of links which both contain (11.03.2024).*

Lastly, we repeat the above experiment but on a more systematic basis. We compare the website references in Copilot's footnotes compared to Bing Search's organic results for multiple products and compute their Jaccard similarity (between the two groups) - See Appendix A for further details. We remove the ads from Bing's Search results for this and any image links. Importantly, we only use queries for which Copilot provides us with Bing Search as one of the footnoted links, because only then we do know the exact corresponding Search query to use for Bing. This means that out of the 109 queried products on Bing Copilot, a www.bing.com search link appears as a footnote website for only 52 of those products.

Figure 5 shows the Jaccard similarity between the two groups of URLs from Bing's Copilot footnotes and Bing Search's organic results for the range of products listed on the vertical y-axis. Out of these 52 products depicted in Figure 5, the results for only two products are the same (though ignoring the exact ordering of results). In half of the 52 products, the Jaccard similarity is equal to or lower than 0.35, which is considered moderately similar. This means that only 35% of the website URLS elements referenced as Copilot sources, or listed in Search, are shared between Copilot and Bing Search. The remaining 65% of the elements are unique to one set or another.

Figure 5. Similarity of Links Shown in Bing Copilot vs Bing Search



Note: *We remove the ads from Bing's Search results for this. We only use queries for which Copilot provide us with Bing Search as one of the footnoted links, because only then we do know the exact corresponding Search query to use for Bing. For normal Bing Search results we go to www.bing.com and we scrape search results by typing the Copilot recommend search query and using Search instead of Copilot. We ignore any LLM results generated at the top. The Jaccard similarity ignores the order of results and just looks at the overlap (the intersection) between the two groups (sets), divided by the size of the union of the two. A value of 1 means full similarity (overlap) and a value of 0 implies no similarity at all. The values of the Jaccard similarity are shown on the x-axis of the bar chart (08.03.2024).*

3 Recommendations

Transparency and algorithmic accountability recommendations covering Microsoft's Copilot would have incredible reach and impact if implemented. Bing's Copilot is already integrated into Microsoft 365 and Microsoft Office, reaching millions of users around the world (Cavanell 2024).

We recommend mandatory, public, disclosures made by AI systems and LLM models – in their 10-K reports or as part of their existing obligations within the EU – on any paid information used for model training, for context window expansion, to help generate an LLM's answers, or to direct user traffic through paid links. This builds on our previous work on disclosures for Big Tech platforms (O'Reilly, Strauss, and Mazzucato Mazzucato 2023; Mariana Mazzucato, Strauss, et al. 2023), including for AI systems (O'Reilly, Strauss, and Mariana Mazzucato 2024). Without more information on AI systems from enhanced disclosures, it is unlikely that antitrust regulations can be configured – especially in an *ex ante* manner – for these markets O'Reilly (2023).

More specifically, immediate transparency in retrieval-augmented generation (RAG) systems is important to mitigate harms (such as biases) from monetization. We, therefore, recommend: listing prominently to the user any paid data sources that the RAG system uses, disclosing publicly when paid factors play a role in the rankings / generated output of any generative AI model, and how ranking factors are weighted in general by core generative AI algorithms (where possible).

In contrast to the "black box" neural network model that drives a foundational model's direct outputs, the ranking mechanisms in the back-end of the RAG system are understandable and so it's ranking factors could be made transparent, explainable, and understandable.

How advertising bidding algorithms work within LLMs and RAGs is vital to disclose to regulators and for regulators to inspect going forward. Many of the harms in Facebook, Google, Amazon and other platforms occurred through advertising algorithms accepting excessively low quality advertising inventory allowing for false, misleading, and dangerous information to spread (Doctorow 2023; CNBC 2023; Bryan 2024). Limiting the spread of knowingly inferior quality paid information through generate AI interfaces is vital given the high degree of trust which consumers are likely to develop in AI systems.

References

Areeda, Phillip and Herbert Hovenkamp (2023). *Fundamentals of Antitrust Law*. Online Edition. Wolters Kluwer.

Baye, Michael R and Jeffrey Prince (2020). “The Economics of Digital Platforms: A guide for regulators.” *The Global Antitrust Institute Report on the Digital Economy* 34.

Bryan, Kenza (2024). “UK advertising watchdog turns to AI to help police ‘dodgy’ green claims.” Online: https://www.ft.com/content/d257eed5-a9ea-4e49-b5ab-663ba9618636.

Cavanell, Zachary (2024). “How to get ready for Copilot for Microsoft 365: Updates for 2024.” Online: https://techcommunity.microsoft.com/t5/microsoft-mechanics-blog/how-to-get-ready-for-copilot-for-microsoft-365-updates-for-2024/ba-p/4037931.

CNBC (2023). “Jeff Bezos urged Amazon to flood search results with junk ads: FTC.” *CNBC*. Online: https://www.cnbc.com/2023/11/02/jeff-bezos-urged-amazon-to-flood-search-results-with-junk-ads-ftc.html.

Doctorow, Cory (2023). *This is your brain on fraud apologetics*. Online: https://doctorow.medium.com/this-is-your-brain-on-fraud-apologetics-59aa5e871c36.

European Commission (2024). “Call for Contributions: Virtual Worlds and Generative AI.” Policy Document. Directorate-General for Competition, European Commission.

Hovenkamp, Erik (2018). “Platform Antitrust.” *Journal of Corporation Law* 44, p. 713.

Jaccard, Paul (1901). “Étude comparative de la distribution florale dans une portion des Alpes et des Jura.” *Bull Soc Vaudoise Sci Nat* 37, pp. 547–579.

Lewis, Patrick et al. (2020). “Retrieval-augmented generation for knowledge-intensive nlp tasks.” *Advances in Neural Information Processing Systems* 33, pp. 9459–9474.

Mazzucato, Mariana, Josh Ryan-Collins, and Giorgos Gouzoulis (2023). “Mapping Modern Economic Rents: The good, the bad, and the grey areas.” *Cambridge Journal of Economics* 47.3, pp. 507–534.

Mazzucato, Mariana and Ilan Strauss (2024). “Algorithm Governance: Five Principles to Prevent AI from Aggravating Digital Economy Harms.” Online: https://www.project-syndicate.org/commentary/algorithm-governance-five-principles-to-prevent-ai-from-aggravating-digital-economy-harms-by-mariana-mazzucato-and-ilan-strauss-2024-02.

Mazzucato, Mariana, Ilan Strauss, et al. (2023). “Regulating Big Tech: the role of enhanced disclosures.” *Oxford Review of Economic Policy* 39.1, pp. 47–69.

McCurley, Kevin S (2007). “Income Inequality in the Attention Economy.” *Google Research*.

O’Reilly, Tim (2023). *You Can’t Regulate What You Don’t Understand*. Online: https://www.oreilly.com/content/you-cant-regulate-what-you-dont-understand-2/.

— (2024). *AI Has an Uber Problem*. Online: https://www.theinformation.com/articles/ai-has-an-uber-problem.

O’Reilly, Tim, Ilan Strauss, and Mariana Mazzucato (2024). “Algorithmic Attention Rents: A theory of digital platform market power.” *Data & Policy* 6, e6.

O’Reilly, Tim, Ilan Strauss, and Mazzucato Mazzucato (2023). “Regulating Big Tech through digital disclosures.” Policy Brief No.26. UCL Institute for Innovation and Public Purpose.

OpenAI (2023). *Retrieval-Augmented Generation (RAG) and Semantic Search for GPTs*. Online: https://help.openai.com/en/articles/8868588-retrieval-augmented-generation-rag-and-semantic-search-for-gpts.

Rock, Rock et al. (2023). “Behind the Clicks: Can Amazon allocate user attention as it pleases?” Working Paper Series IIPP WP 2023-11. UCL Institute for Innovation and Public Purpose. Online: https://www.ucl.ac.uk/bartlett/public-purpose/wp2023-11.

Roth, Alvin E (2015). *Who Gets What–and Why: The New Economics of Matchmaking and Market Design*. Houghton Mifflin Harcourt.

Strauss, Ilan, Tim O'Reilly, and Mariana Mazzucato (2024). "Amazon's Algorithmic Rents: The economics of information on Amazon." *UC Law Science and Technology Journal*. Forthcoming spring 2024.

Ursu, Raluca M (2018). "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions." *Marketing Science* 37.4, pp. 530–552.

Wikipedia contributors (2024). *Jaccard index – Wikipedia, The Free Encyclopedia*. [Online; accessed 11-March-2024]. Online: https://en.wikipedia.org/wiki/Jaccard_index.

A Appendix: Bing Copilot Methodology and Testing

A.1 Bing Copilot: Repeated Prompts

We initiated the study by examining the sequence of suggested URLs offered by Bing Copilot in response to the query "Where should I buy a Bose Quietcomfort 45?". We then compared these suggestions with the (organic) URLs recommended by the Bing Search engine, which were generated based on the terms proposed by Bing Copilot. In both cases we used a US IP address and noted that the results differ significantly to the results that would be shown to a user with a German or Swiss IP address. Moreover, while this research has been conducted Microsoft changed the product naming from Bing Chat to Bing Copilot and likely some underlying architecture features. This data was collected 11 March 2024.

A.2 Bing Copilot: Main Case Study

We asked ChatGPT for the top 100 electronics products ("Can you give me a Python list of top 100 electronic products?") which resulted in 109 products. The code from the repository here was adapted to query Bing Copilot about these 109 products. The specific query used was "Where should I buy a {product}?" where the "product" term is substituted with each of the electronic products for the query.

Where the suggested footnote links by Bing Copilot contained a reference to Bing Search, that same query was used to query Bing Search. The code built on the Search-Engines-Scraper repository that can be found here. The queries are similar to this form: "Bose+Sport+Earbuds+buy+online". The suggested links by Bing were collected noting that these do not include the advertisement links. Bing Copilot has a limit of 30 requests before throttling. In addition, requests were made using an US IP relying on Proton's VPN service. The data was collected between 19 February 2024 and 1 March 2024.

As a next step, the suggested link lists from Bing Copilot and Bing Search (where applicable) are trimmed to the size of the smaller list to enable comparison. The order was preserved. Getty image links have been removed from the lists as they refer to the images that are sometimes included or added to the text.

These two lists are then compared using the Jaccard Similarity score.

The Jaccard similarity score is a measure of similarity between two sets. It is defined as the size of the intersection of the sets divided by the size of the union of the sets. In mathematical terms, the Jaccard similarity score $J(A, B)$ between two sets A and B is given by (Jaccard 1901; Wikipedia contributors

2024):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Where:

- $|A|$ represents the cardinality (number of elements) of set A.
- $|B|$ represents the cardinality (number of elements) of set B.
- $|A \cap B|$ represents the size of the intersection of sets A and B, i.e., the number of elements that are common to both sets.
- $|A \cup B|$ represents the size of the union of sets A and B, i.e., the total number of unique elements in both sets.

The Jaccard similarity score ranges from 0 to 1, where 0 indicates no similarity (no common elements between the sets) and 1 indicates perfect similarity (both sets are identical).

A.3 Additional testing

We wanted to see whether there was a correlation between a product's price and the nature of the Bing query link in the footnote. For electronics products mentioned before, we could not identify a significant difference between the mean price for those two groups (Student's t-test).

We retrieved the 30 bestselling headphone products and books from Amazon and queried Bing Copilot for the exact product title on Amazon's U.S. website. We found that in both product categories more than 80% of the Bing Copilot results do not contain reference footnotes to Bing Search. The data was collected between 19 February 2024 and 8 March 2024.